Volume Title: Measuring Entrepreneurial Businesses: Current Knowledge and Challenges

Volume Author/Editor: John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-45407-8 (cloth); 978-0-226-45410-8 (e-ISBN)

Volume URL: http://www.nber.org/books/halt14-1

Conference Date: December 16-17, 2014

Publication Date: September 2017

Chapter Title: Immigrant Entrepreneurship

Chapter Author(s): Sari Pekkala Kerr, William R. Kerr

Chapter URL: http://www.nber.org/chapters/c13502

Chapter pages in book: (p. 187 – 249)

5

Immigrant Entrepreneurship

Sari Pekkala Kerr and William R. Kerr

5.1 Introduction

Immigrant entrepreneurship is of central policy interest and a frequent hot point in the popular press. Many policymakers believe that immigrant founders are an important and underutilized lever for the revival of US job growth and continued recovery from the Great Recession. Several local and national policy initiatives have been launched to attract immigrant entrepreneurs (e.g., the Thrive competition in New York City, the Office of New Americans in Chicago, and the White House Startup America initiative). Some of the policy initiatives focus on specific issues that have been found to limit immigrant entrepreneurs from starting or growing their businesses (e.g., language barriers, difficulty navigating the legal steps to start a company, or lack of capital to pilot projects), while others are generally focused on attracting more new businesses into the country. Policies vary

Sari Pekkala Kerr is an economist and senior research scientist at the Wellesley Centers for Women (WCW) at Wellesley College. William R. Kerr is the Dimitri V. D'Arbeloff-MBA Class of 1955 Professor of Business Administration at Harvard Business School and a research associate of the National Bureau of Economic Research.

in the immigrant group that they target, ranging from a specific focus on high-skilled immigrant entrepreneurship with venture capital (VC) backing to broad-based measures that potentially touch many diverse immigrant communities.

Academic research, unfortunately, possesses only a small voice in this debate or policy design. For example, advocates of greater immigrant entrepreneurship mainly cite a few extreme examples of success such as Sergey Brin, one of the founders of Google, and extrapolate information from some exceptionally influential case studies regarding Silicon Valley and large high-tech companies (e.g., Saxenian 1999; Wadhwa et al. 2007). While each of these supporting pieces has its merits and liabilities, it is important for rigorous trends and statistics to also inform this debate. For example, even with respect to Silicon Valley and high-tech companies, it is not immediately clear what the oft-cited statistics mean—it is likely true that more than half of the entrepreneurs in Silicon Valley are of immigrant origin, as the surveys suggest, but the exact same could be said of the undergraduate student populations at most schools in the San Francisco area. Second, given the substantial heterogeneity in immigrant entrepreneurship, which comes in just as many flavors as native entrepreneurship, it is unsatisfying to focus on such a narrow population of high-tech entrepreneurs for contemplating possible initiatives.

For economists to be able to aid the policy process, and ultimately improve economic performance in this arena, we need better-grounded estimates on the importance of immigrant groups for the creation of new firms, the business activities and growth profiles of created firms, and so on. This study constructs a data platform using Census Bureau administrative data to assist in this process. The purpose of this chapter is to detail the platform's components and provide some early views of the trends for immigrant entrepreneurs and the patterns in their behavior. We have several audiences in mind. First, we are able to offer some new facts to the discussion of immigrant entrepreneurship that can be useful for policy discussions, although we do not examine any specific policy actions or recommendations about encouraging immigrant entrepreneurship in this study. Second, we hope that others find this discussion encouraging for making progress on this front and that they too seek access to these data through the Census Bureau. Ideally, our chapter can provide them a one-stop-shop for what is feasible in the data and how to build the platform, and this chapter goes into greater depth than is normal for academic papers on how the platform is built and its traits. Finally, we speak to future efforts to enhance these data. In terms of representative statistics, this platform is likely as good as it gets with today's data collection. We describe below a wish list for future data development efforts.

Section 5.2 provides a brief review of the previous empirical literature on immigrant entrepreneurs and their traits. Fairlie and Lofstrom (2014) provide a comprehensive recent review of this literature strand and statistics

from the 2007 Survey of Business Owners (SBO). In this chapter, relative to prior academic work, we make three contributions. First, our data platform provides consistent estimates of immigrant entrepreneurship over a long time period (1995–2008) and across skill levels (e.g., all entrepreneurs, VC-backed firms). Existing work, even when using representative national samples, tends to be cross-sectional at a given point in time and focused on a specific skill population, whereas for most purposes the comparisons across time and groups would be very important. Second, we study the different dynamics of employment and growth among immigrant-led businesses compared to those founded by natives. Fairlie and Lofstrom (2014) conclude that a central research goal for immigrant entrepreneurship is to identify the dynamics of employment growth among these firms, and our constructed platform makes progress in this domain. Third and related, we provide a first breakdown of these growth dynamics by the age of immigration to the United States. This last analysis is preliminary and mostly undertaken to show the potential of the data for observing differences along traits identifiable from the 2000 Census match, but striking nonetheless.

Section 5.3 details the construction of our data platform. The strength of our study lies in the ability to use and combine several restricted-access US Census Bureau data sets to create a unique longitudinal data platform with millions of observations. Indeed, a key purpose of this chapter is to report on the potential of these data and describe their traits for research purposes. The backbone for our work is the employer-employee data in the Longitudinal Employer-Household Dynamics (LEHD) database. The LEHD provides firm-worker information collected from unemployment insurance records. From this information, one can observe the birth of new firms and their employee composition, including immigration status. We also utilize other data: (a) the Longitudinal Business Database (LBD) to assess the employment growth dynamics of new firms, (b) the long-form records of the 2000 Decennial Census to collect additional person- and household-level traits (e.g., year of arrival to the United States), and (c) external data sets to identify companies that have VC backing or achieve an initial public offering (IPO). The resulting data platform can describe many forms of entrepreneurship, ranging from general patterns to "growth entrepreneurs" described in the VC literature. These detailed new data allow us to study person- and firm-level patterns in a way that has not been possible to date, and this section also depicts the variations and limits on the definitions of entrepreneurship in the LEHD.[1]

Section 5.4 provides our trend estimates. Most of our work focuses on eleven states present in the LEHD since 1992, which include California and

---

1. In a broad review of the immigration literature, Lewis (2013b) raises immigrant entrepreneurship as one of the key areas requiring further study and notes the key ingredient of employer-employee data for this purpose. The LEHD is the source of this type of administrative information for the United States.

Florida. We estimate 24 percent of entrepreneurs in these states from 1995 to 2008 are immigrants, which broadly corresponds with other studies. As important, this immigrant entrepreneurship share rises from 17 percent in 1995 to 27 percent in 2008. Our sample, by coincidence, draws from heavy immigration states. Looking at a sample of twenty-eight states present in the LEHD by 2000, we estimate these numbers may be 3 percent higher than the total population, with the growth trend being similar. Returning to our focal set of states, the immigrant share among entrepreneurs receiving VC financing is modestly higher, reaching 30 percent in 2005 compared to 27 percent for all firms. In terms of entrepreneurship rates, roughly 2 percent of immigrants start a business over a three-year period; 0.1 percent start a firm backed by VC financing. These rates are higher than those we estimate for natives, which is reflected in the fact that immigrants constitute 19 percent of the LEHD workforce in our sample, less than the entrepreneurship shares reported above.

Section 5.5 documents basic patterns related to entry, survival, and growth of immigrant versus native businesses. On the whole, the businesses started by immigrant entrepreneurs perform better than native businesses with respect to employment growth over three- and six-year horizons. This is evident in the raw data and mostly persists when comparing immigrant- and native-founded businesses started in the same city, industry, and year. By contrast, immigrant-founded businesses show no advantages with respect to payroll growth, and may in fact generate lower-wage job growth. Combining business survival with growth dynamics, immigrants tend to be engaged in more volatile, up-or-out type dynamics, along the lines described by Haltiwanger, Jarmin, and Miranda (2013) for young businesses and job creation. Most of this effect is captured by the city-industry-year choices made for businesses, versus variation in growth patterns within these cells. Breaking down these aggregate results, the strongest employment growth impacts for immigrants are found in high-wage businesses and high-tech sectors. Conditional on receiving VC investment, however, we do not observe greater business survival, better employment outcomes, or higher likelihood of going public for firms founded by immigrant entrepreneurs. Businesses founded by immigrants who came to the United States by age eighteen have stronger growth patterns than those founded by immigrants migrating as adults.

Section 5.6 concludes with a discussion of these findings in the larger context of immigration and entrepreneurship policy. This section also describes enhancements to the platform that would enable better research efforts going forward. Immigration policy is designed and administered at the national level, with few restrictions on the location choices of immigrants within the United States. A methodological conclusion from this study is that the LEHD-based platform can consistently describe immigrant entrepreneurship across many industries, geographies, and skill levels. This is an important ingredient for delivering systematic advice about how

immigrant entrepreneurs impact the US economy and where the effects of expansion in admission levels might be felt. We also find that the detail of the LEHD can aid more effective policy advice for subgroups of immigrants. The surveys most cited in the public discourse are actually quite crude in this regard. For example, although Sergey Brin is often used as the showcase example for immigration reform on high-skilled H-1B policies, he came to the United States as a child, while the H-1B program is focused on temporary admissions of adults with college-level education. Our disaggregation by the age at arrival for immigrants indicates this assembled platform can help begin to disentangle these important details. To be clear, our study stops well short of making a direct input to immigration policy design on this or other dimensions, but we do find that these data elements are of sufficient strength and depth that they can serve as an effective foundation for future research efforts to inform the economic consequences of selection strategies based upon certain immigrant traits.

On the other hand, we note that there are many places where the LEHD has limits, some of which may be addressable. For example, here we define entrepreneurship status through initial wage profiles of firm employees, which is certainly incomplete. Over short-term horizons, it would be useful to consider linking LEHD workers to SBO-type data to evaluate the accuracy and bias of derived entrepreneurship definitions with greater detail. Similarly, we describe how links of LEHD workers to external data sources at the individual level would be powerful (e.g., entrepreneurs/CEOs contained in the Venture Xpert data, inventors contained in United States patent data). More challenging, while we are able to make progress toward some of the important traits of immigrants, we miss completely with the LEHD the essential questions of immigration visa type (e.g., H-1B, green card), which is strongly emphasized by Hunt (2011) as a key predictor for choices by immigrants with respect to entrepreneurship. For the evaluation and design of effective policy, the inclusion of visa status and transitions over time must be at the top of any wish list.

5.2   Previous Literature on Immigrant Entrepreneurs

This section provides some background for our study from the academic literature, with Fairlie and Lofstrom (2014) reviewing the immigrant entrepreneurship literature in a more comprehensive manner. There is a large body of literature showing that general rates of business ownership are higher among the foreign born than natives in many developed countries, including the United States, United Kingdom, Canada, and Australia.[2] Fairlie (2012) and Fairlie and Lofstrom (2014) also observe that trends in

---

2. Studies include Borjas (1986), Lofstrom (2002), Clark and Drinkwater (2000, 2006), Schuetze and Antecol (2007), and Fairlie, Zissimopoulos, and Krashinsky (2010).

self-employment rates and new business formation are increasing among immigrants but decreasing among natives in the United States. In closely related work to the current analysis, Hunt (2011, 2015) focuses on skilled immigrants and finds that they are more likely to start firms with more than ten employees than comparable natives. Hunt uses the National Survey of College Graduates (NSCG), which provides a nationally representative sample of persons with a college degree and interesting information about the initial visa status of immigrants. She finds that the probability of starting a firm was highest for those initially arriving on a study visa or a work visa versus family reunification. While the level of detail on the specific degrees and entry visas in the NSCG is impressive, the major issues for researchers trying to describe national trends in immigrant-founded firms are the small sample sizes, the lack of longitudinal dimension, and the absence of firms as a data element.

In parallel to these general patterns, a second research stream focuses specifically on immigrant entrepreneurship in the high-tech sector. Saxenian (1999, 2002) documents that up to a quarter of the high-tech firms in Silicon Valley in the 1980s and 1990s were founded or being run by immigrants. Wadhwa et al. (2007) extends this survey analysis to the rest of the country and other industries to study firms founded in 1995–2005. They document similar shares of immigrant-founded companies across the country, although elements of their study are difficult to generalize.[3] Table 5.1 provides a summary of several related studies on immigrant entrepreneurs. These efforts connect into a broader line of work showing the overrepresentation of skilled immigrants among certain extreme outcomes in US science and engineering: for example, US-based Nobel Prize winners (Peri 2007), high-impact companies (Hart and Acs 2011), patent applications (Wadhwa et al. 2007), members of the National Academy of Sciences and the National Academy of Engineering (Stephan and Levin 2001), and biotech companies undergoing initial public offerings (Stephan and Levin 2001). It is these kinds of factors that likely drive the current policy enthusiasm toward immigrant entrepreneurs. On the other hand, with respect to immigrant high-tech contributions, Hart and Acs (2011) suggest that although immigrants play an important role, "most previous studies have overstated the role of immigrants in high-tech entrepreneurship."

---

3. The sample with responses covers 7 percent of the approximately 30,000 firms in the Dunn & Bradstreet data that were founded in 1995–2005, had sales greater than $1 million, and employed at least fifty persons. Despite the extensive efforts of their research team to reach a subset of companies listed in the Dunn & Bradstreet data, the study faces a lower response rate, selectivity in terms of which firms choose to respond, and perhaps limits regarding the ability of the surveys to reach the right person to answer the questions related to the founders' origin (as an HR or a PR person might not know whether one of the founders moved to the United States as a child). Therefore, the researchers extrapolate from their sample to produce nationally representative numbers for revenue and employment generation. Monti, Smith-Doerr, and MacQuaid (2007) provide related evidence from Massachusetts high-tech firms.

Table 5.1 Previous studies on immigrant entrepreneurship

Study (1)	Data sources (2)	Sample design (3)	Period (4)	Immigrant entrepreneur share (%) (5)	Founding rate among immigrants (%) (6)	Founding rate among natives (%) (7)
Saxenian (1999)	Dun & Bradstreet, surveys	Silicon Valley tech firms	1980–1998	24	n/a	n/a
Anderson & Platzer (2006)	Thomson Financial, survey, Internet	340 publicly traded, venture capital-backed firms, independent in 1990–2005	1990–2005	25	n/a	n/a
Wadhwa et al. (2007)	Dun & Bradstreet, surveys	US tech firms with > $1m sales and > 20 employees	1995–2005	25	n/a	n/a
Reynolds & Curtin (2007)	Panel Study of Entrepreneurial Dynamics	New founders who plan to grow firm to 50+ employees	1999	15	n/a	n/a
Monti, Smith-Doerr, and MacQuaid (2007)	MA Biotech Council member list	New England biotech firms	2006	26	n/a	n/a
Fairlie (2008)	Census IPUMS 2000, Current Population Survey 1996–2007	Working-age business owners, work > 15 hours per week	2000, 1996–2007	17	0.4	0.3
Hart & Acs (2011)	Survey	"High-impact" high-tech companies	2002–2006	16	n/a	n/a
Robb et al. (2010)	Kauffman Firm Survey	High- and medium-tech firms founded in 2004	2004	10.3	n/a	n/a
		Firms founded in 2004 that survived until to 2008	2008	9.4	n/a	n/a
Hunt (2011)	NSCG 2003	College degree holders, working in 2003	1998–2003	n/a	0.8	0.6
Fairlie (2012)	Survey of Business Owners 2007, Current Population Survey 1996–2010	Owners of businesses in 2007, workforce in 2010	2007, 2010	13	0.6	0.3

Notes: The immigrant share of entrepreneurs in Saxenian (1999) covers only Chinese and Indian ethnic groups, based on the CEO surnames. The founding rate in Hunt (2011) covers firms founded between 1988 and 2003 that have at least ten employees in 2003. The founding rates in Fairlie (2008, 2012) are monthly. Fairlie (2008) defines the immigrant entrepreneur share as the percentage of new companies that had at least one immigrant founder (of all business owners, 12.5 percent are immigrants). Hart and Acs (2011) use Corporate Research Board's American Corporate Statistical Library (ACSL) listings to identify the set of "high-impact" companies and obtain survey responses from 29 percent of them.

Statistics with respect to immigrant entrepreneurship among VC-backed firms are harder to assemble. Fairlie (2012) calculates from the 2007 SBO that equally few native and immigrant business owners rely on any VC funding; more generally, the study finds that immigrants are less likely to start a business with no capital and just as likely as natives to start a business with more than $1 million of capital. Hegde and Tumlinson (2014) and Bengtsson and Hsu (2015) are recent examples of studies of immigrant entrepreneurship among VC-backed firms using ethnic names to distinguish the likelihood of a founder being an immigrant. In an advocacy piece, Anderson and Platzer (2006) identify the higher immigrant entrepreneur share among publicly listed VC-backed US companies.

Related to our focus on entrepreneurs and how immigration influences the supply of these individuals, prior studies show that more educated business owners run more successful businesses, generate more innovations, and grow their firms faster over time (e.g., Unger et al. 2011). There is an overall positive relationship between education and business ownership, although the evidence is somewhat mixed as highlighted in the meta-analysis of van der Sluis, van Praag, and Vijverberg (2008). Lofstrom, Bates, and Parker (2014) postulate that this may be due to sorting into industries based on entry barriers. They find that educational credentials of highly educated potential entrepreneurs are associated with lower probability of small-firm ownership in less financially rewarding industries, while they increase entry into higher-barrier industries offering higher returns.

Immigrants can take a variety of paths into firm ownership in the United States. Many skilled immigrants enter the United States for study or paid work and found their company after several years in the country, while a smaller number enter for the specific purpose of opening a business. Kerr, Kerr, and Lincoln (2015a, 2015b) describe in greater detail how US immigration law and corporate sponsorship of visas contribute to this career trajectory. Akee, Jaeger, and Tatsiramos (2007) find that premigration self-employment in home countries increases the probability of self-employment by immigrants in the United States and boosts self-employment earnings. Lofstrom (2002) finds that self-employment probabilities and earnings for immigrants increase with time spent in the United States, perhaps even reaching earnings parity with observationally similar US-born entrepreneurs after about twenty-five years in the country. The use of repeated cross sections of censuses, however, limits the degree to which the role of assimilation can be separated from selective out-migration. Blau and Kahn (2016) describe cultural factors influencing gender-based rates of assimilation for work by immigrants.

The spillover effects to native opportunities for opening a business have been examined by several research teams. Fairlie and Meyer (2003) find some evidence that immigration may negatively affect native self-employment probabilities using the 1980 and 1990 Census records. Other researchers

suggest skilled immigrants generate positive spillover effects in local areas. For example, recent work points toward positive spillover effects for cities or states when measured in terms of innovation, publications, and productivity (e.g., Kerr and Lincoln 2010; Hunt and Gauthier-Loiselle 2010; Peri, Shih, and Sparber 2015). Lewis and Peri (2015) provide a theoretical framework and review of literature on the effects of immigration on local areas. Evidence from more historical contexts is mixed (e.g., Borjas and Doran 2012; Moser, Voena, and Waldinger 2014). Kerr (2013) provides an extended review of literature on skilled immigration and notes the particular gap around the empirics of immigrant entrepreneurship specifically.

Another line of work documents how immigrant entrepreneurs appear to specialize in a narrower range of industries or occupations compared to native entrepreneurs. Very common examples from the United States include Korean entrepreneurs for dry cleaners, Vietnamese entrepreneurs for nail care salons, Gujarati Indian entrepreneurs for the motel industry, and Punjabi Indian entrepreneurs for convenience stores. Chung and Kalnins (2006) provide a first analysis of this specialization for US Indian entrepreneurs, and Patel and Vella (2013) document patterns more broadly and show earning consequences. Fairlie, Zissimopoulos, and Krashinsky (2010) provide cross-country comparisons for some groups. Kerr and Mandorff (2015) provide a theoretical model and empirical evidence of how small group sizes and social isolation can provide comparative advantages for ethnicities in self-employed sectors where the entrepreneurs benefit from the tight networking of their social group.

Compared to this earlier literature, the LEHD-based platform has the capacity to provide critical and novel information for the enhanced understanding of immigrant entrepreneurship and the effective calibration of immigration policy. Indeed, even though earlier work has tackled many important issues, there remain unfortunate gaps in both the big-picture study of immigrant entrepreneurship and in the depth of insights feasible.

Starting with the big picture, immigrant entrepreneurship is often greatly hyped in both policy and media circles, and a number of newspaper and business press articles (e.g., *Forbes*, *New York Times*) tout immigrant founders as the solution to the sluggish recovery from the Great Recession. Similarly, current local and national policy efforts seek to attract and support new firm formation by immigrants, as noted in the introduction. Many of the studies in table 5.1 are limited by sample designs that are not broadly representative of the economy, or they are cross-sectional in nature if representative. This creates a credibility challenge for the work, even when undertaken with the utmost objectivity; the gap gets extremely large with advocacy pieces or those concentrating only on the most prominent high-tech clusters. Building directly upon administrative data, an LEHD-based platform can provide a substantially stronger foundation for its covered states and the credibility necessary for grounding debates around common facts.

The depth is also essential. Some studies, such as Fairlie (2008, 2012) and Fairlie et al. (2015), provide attractive estimates from the Current Population Survey (CPS) that are the current best practice for longitudinal series. The CPS is in many respects a solid platform, given its stable data collection and long series, and we compare some of our work to the CPS later. The difficulty with the CPS is the ultimate depth that it can provide. Its sample sizes of about 500,000 adults are large enough to provide annual estimates for states or industries, but the cell counts become too small when attempting to jointly view these traits. Moreover, the CPS records cannot be linked to future growth trajectories of the firms, the use of venture-capital funding, and so on. The CPS relies on founders declaring themselves to be self-employed and yet also does not measure if other employment is being generated. Thus, while the CPS is an important and publicly available resource, the development of an LEHD-based platform that includes every business in covered states will provide much deeper capacity for statistical and analytical work. As we describe below, there is no question that the LEHD is far from perfect in terms of what it can do; on the other hand, if these limitations are acceptable, then the scope of follow-on work becomes extraordinary. This depth is true in terms of potentially publishable statistics of entry rates, and also in terms of more complex academic questions (e.g., how many prior employers do immigrants typically have before starting a firm and how does this prior job history impact entrant performance).

Reading the anecdotal accounts and collected statistics regarding immigrants being very involved in high-growth entrepreneurship, it seems natural from a policy perspective to want to encourage this development. Encouraging immigrant entrepreneurs seems like an essential prong of science policy, and its mandate reaches the highest levels of government. Yet, the earlier research has developed only partial data needed to effectively evaluate these policies or enhance their design. Research based upon case studies or small surveys may identify the trend, but they fail to build the empirical foundation necessary for confidence in the design of proposed legislation and the likely impact of reforms. The data platform introduced here makes significant headway in that respect as it will utilize and combine universal and unique microdata sets on individuals and firms (LEHD, LBD, and the Decennial Censuses). In short, this provides a unique platform for the study of immigration and entrepreneurship.

5.3   Census Bureau Data and Measurement

5.3.1   Data Platform

The LEHD is the centerpiece of our platform, in combination with the LBD and a person-level match to the 2000 Decennial Census of Population (census). These data sets are confidential, housed by the US Census Bureau,

and accessible via research data centers. Built from quarterly worker-level filings by employers for the administration of state unemployment insurance (UI) benefit programs, the LEHD identifies the employees of each private-sector firm in the United States and their quarterly compensation. It is longitudinally linked at both the firm and employee levels, allowing one to model how firm employment structures adjust over time, how new entrepreneurial firms form, and how individuals transition from wage work into entrepreneurship. This rich data source is currently available for thirty-one states for research purposes, and it will eventually cover the whole country.

The initial LEHD dates vary by state, and we focus on two samples in this chapter. The first is the eleven states that have LEHD records that begin by 1992 or earlier: California, Colorado, Florida, Idaho, Illinois, Louisiana, Maryland, North Carolina, Oregon, Washington, and Wisconsin. Our data extend through 2008, and the newest vintage of the LEHD continues coverage until 2011. Certain analyses are also conducted on a larger set of twenty-eight states present in the LEHD by 2000, as shown in appendix table 5A.1 and figure 5.1. The larger sample can only be followed over a short time span, but it helps us understand the impact of state-level variation in immigration and entrepreneurship rates for the trends presented in this chapter, particularly the inclusion of immigrant-heavy California and Florida among our primary sample.

The LEHD Individual Characteristics File (ICF) contains basic information about individuals such as age, gender, race, place of birth, and citizenship status. Through the Employment History Files (EHF), one can also discern earnings and employment histories of each person job-by-job and in aggregate. In addition, unique person identifiers (PIKs) afford matches of the LEHD to the individual-level records contained in the 2000 Census of Populations. The PIKs are anonymous identifiers that match one-for-one with Social Security numbers. The Census long-form responses cover one-sixth of the US population, allowing us to link census responses for roughly 16 percent of our LEHD workers. From the 2000 Census, we extract individual-level characteristics from the Person File and household and housing-unit characteristics from the Household File. Long-form responses contain very detailed person and household characteristics (e.g., year of entry into the United States, level of education, occupation, marital status, family composition, household income by source, etc.). It is important to recognize that while the LEHD covers employees from the early 1990s through 2008—including new immigrants at the end of the end of the sample period—the 2000 Census match requires individuals to be living in the United States by 2000.

We build a tailored data set for the analysis of immigrant status and entry into entrepreneurship, first focusing on the dynamics over time. We take several steps to reduce the set of massive data records into a manageable platform that properly represents the phenomenon of interest, and it
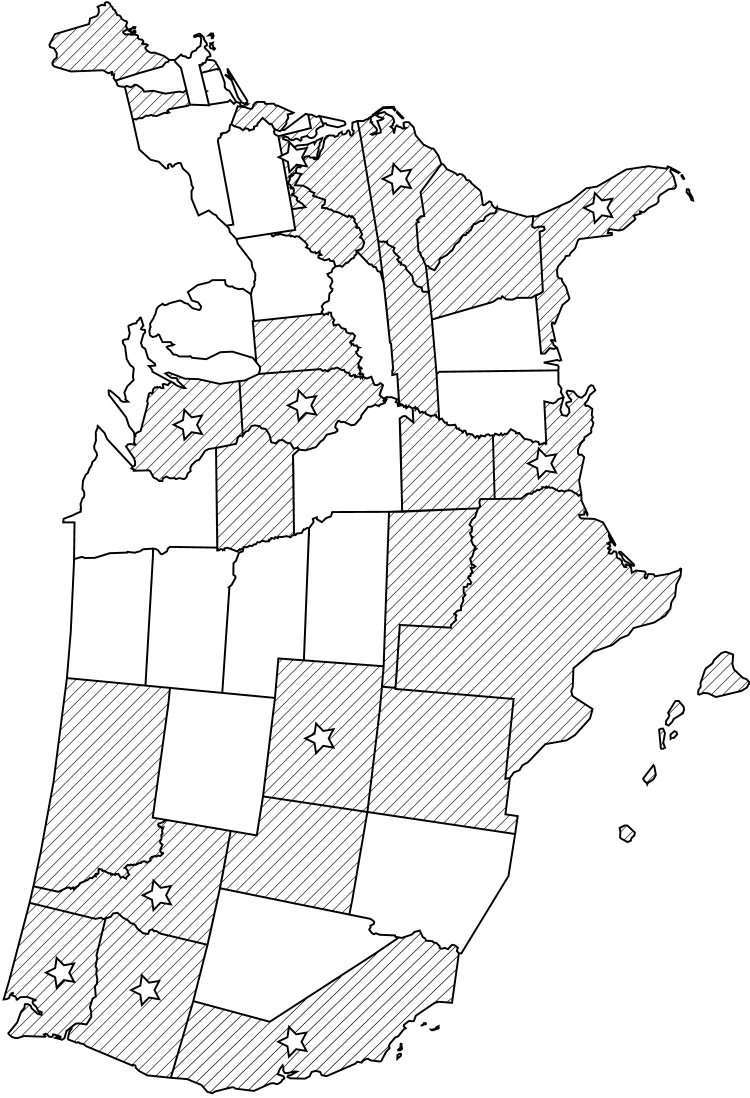
Fig. 5.1 LEHD state coverage

Notes: The figure indicates with shading the states that are covered by the 2008 version of the LEHD used in this study. Alaska is not covered. Stars indicate the eleven states whose coverage begins by 1992. The wider sample used in this study includes all shaded states excepting Arkansas. Coverage for all states ends in 2008.

is important to describe these steps as they have some bearing on our measurements. We begin by retaining for each individual his or her main job during the year (i.e., the job from which the person obtains the majority of their LEHD earnings). We also only retain persons for whom the average quarterly earnings from the main job are at least $2,000 per quarter. We further restrict our sample in each year to individuals age twenty-five to fifty. This age restriction is such that we stay reasonably far away from retirement decisions—and, in particular, the formation of small-scale businesses as a form of semiretirement—and concentrate on entrepreneurial activity in the peak employment years of each person's working life. Moreover, we require individuals be present in the LEHD at least three quarters that span two calendar years.

Immigrants are simply identified as those persons born outside of the United States. This information is available for all LEHD individuals and is based on the Social Security Administration (SSA) Statistical Administrative Records System (i.e., StARS database). Some of these immigrants may have later been naturalized and become citizens, but that information is not utilized in this study. This is partly due to our focus and also due to uncertainty about the updating procedures for this information. By defining immigration status through the ICF files, we can depict immigrant entrepreneurship consistently over the sample period, including new arrivals. We separately consider information from the match with the 2000 Census, which records the year when the immigrant arrived into the United States.

5.3.2 Defining Firm Entry and Entrepreneurs

Our evaluation of entry into entrepreneurship also utilizes the LBD, a business registry that contains annual employment and payroll for all private-sector establishments in the United States since 1976. The LBD and the LEHD use several levels of establishment and firm identifiers: (a) State Employer Identification Numbers (SEIN), (b) federal Employer Identification Numbers (EIN), and (c) the overall company identifier (ALPHA) that the Census Bureau uses to link the establishments of multiunit companies together. Following Haltiwanger, Jarmin, and Miranda (2013), we identify for each establishment the first year during which the firm that the establishment belongs to was observed to be in operation within the LBD. We also create for each firm the number of employees that the LBD reports were working for this firm in the initial year.[4] Approaching entrant definition in this way accomplishes several things—it builds off of the national LBD database to avoid issues related to the partial LEHD state coverage, connects SEINs as appropriate into parent firms, and ensures a consistent

---

4. The data structure of the LEHD and LBD allow for establishments within each firm to have different industries and locations. In rare cases where required in this study, we define the main industry and main location of a multiunit firm through the facility with the largest number of employees.

definition of entry with prior academic work using the Census Bureau data. As two examples, this approach ensures our entrepreneurship definition does not include the formation of a new SEIN by an existing multiunit firm expanding into an LEHD state, nor the development of new SEINs for tax purposes by existing businesses. With respect to the broader literature, our approach focuses on the formation of employer establishments, whereas the commencement of Schedule C self-employment activity is unmeasured and not considered to be entrepreneurship in this sample.[5]

A very important issue, and the weakest link of these data for the study of entrepreneurship, is that the LEHD does not designate the founders or owners of firms. Similarly, compensation data includes bonus pay but not equity ownership of individuals. We use the term "entrepreneur" to describe anyone present in the data who satisfies three criteria: (a) in an entering single-unit firm per the Haltiwanger, Jarmin, and Miranda (2013) definition, (b) present in the LEHD in the first year that the new firm entered, and (c) among the top three initial earners in the firm. With respect to the first condition, we require the new firm be a single-unit firm in its start year to ensure that we have complete employment records from the LEHD. By itself, the second condition focuses on the initial employees of the firm and will in many cases include early hires. The third condition then associates entrepreneurship with the top initial earners in the firm. This will clearly be inaccurate in some cases, and some entrepreneurs deliberately take low salaries or no compensation from their firms early on to conserve funds. On the other hand, as we describe below, most firms in our sample are small and are likely reasonably well modeled by this approach. We also show results that drop the third condition and thereby provide statistics related to all initial employees in the firm.

Terms like "entrepreneurship" are also vague with respect to the time dimension. For example, a strict application of our three-part definition would declare the founder of a new firm to be a wage worker starting in the second year of a firm's life, yet many still consider Mark Zuckerberg to be an entrepreneur a decade after the founding of Facebook. For most of our trend statistics, we accordingly use a three-year window that declares an individual to be an entrepreneur if the firm was founded in the prior three years. We still require that the individual had been present in the year of founding and a top initial earner, the second and third conditions. In fact, we do not require the individual remain necessarily associated with the firm, simply that the firm and individual persist. We present results that alternatively use a strict one-year definition. Overall, an unfortunate trade-off exists in that longer windows for keeping track of entrepreneurs result in shorter sample durations, due to the greater number of preperiod years that must

5. There is scope for further work on this regard using the Integrated LBD that contains nonemployer firms.

be devoted to determining the initial values. Said differently, if we wanted to declare individuals to be entrepreneurs if they have founded a business over the prior ten years, the earliest start date for the LEHD-based sample series is 2001, after the 1990s high-tech boom period that is so interesting to study.

5.3.3   Benchmarks for Definitions

We can use the public-use Survey of Business Owners (SBO) data from 2007 to help benchmark these choices with respect to the top three earners and their limits. The file contains over two million observations on employer and nonemployer firms, and the data contain detailed information about the firm and its owners. We focus on employer firms that mirror the LEHD-based sample built upon UI data, and we drop a small number of firms where the firm does not report whether the first-listed owner is a native or immigrant or no ownership data are present at all. Throughout, we use sample weights to provide population-level statistics.[6]

We first analyze the likelihood that the business represents the owner's primary source of personal income. For the full cross-section of single-owner businesses in 2007, this is true for 81.8 percent of businesses with an immigrant owner and 81.4 percent of businesses with a native owner. Similarly, when incorporating businesses with multiple owners, at least one owner reports the business as the primary income source in 81.9 percent of firms where an immigrant owner is present, very similar to the 81.3 percent rate in firms where no immigrant owner is present. When looking at the most recent entrants (i.e., business founded in 2007), the overall fraction of businesses being the primary income source unsurprisingly declines due to transition issues, but remains at two-thirds. It also tilts modestly toward immigrants— across all 2007 entrants, 70.3 percent and 64.7 percent of immigrant- and native-owned entering businesses constitute the primary ownership source, respectively, and this difference is statistically significant (t-stat = 4.03). Nonetheless, these general patterns are supportive of the use of LEHD records for identifying entrepreneurs, compared to an environment where most owners only derived very modest sums from businesses (e.g., businesses set up for tax purposes, hobby entrepreneurship, or to employ household workers).

We next consider calculations that help evaluate our focus on up to three owners. The news is again mostly supportive. Across all SBO firms in 2007, the average number of owners is 1.8, while for the newest 2007 entrants it is 1.7 owners. With our approach in the LEHD data, the average owner

---

6. This is the first-ever SBO Public Use Microdata Sample and it allows researchers to create their own tabulations and analyses on entrepreneurial activity, including the relationships between firm characteristics such as sources of capital, number of owners, firm size, and firm age. Going forward, a main data objective is to unite our LEHD platform with the confidential version of the SBO. (Data and descriptions are available at: https://www.census.gov/econ /sbo/pums.html.)

number is 2.16. This difference comes mostly from underestimating the share of firms with a single founder, and in some cases we are including an extra person in some of our assignments. On the other hand, a nontrivial share of SBO businesses report four or more owners compared to our cap at three entrepreneurs. Without more information, we must draw the line at some founder count, and we believe three founders provides the best balance.

This trade-off suggests that we need to examine closely how immigrant owners are distributed over the owner count distribution to see if scope for bias exists. Focusing on the 2007 entrants, we find relatively uniform rates of an immigrant owner being present: 23.4 percent, 22.4 percent, 27.0 percent, and 27.3 percent for firms with one, two, three, and four or more owners, respectively. A rising share on this dimension is to be expected, given our focus on any owner being an immigrant, and the differences are very modest. This suggests a rather small scope for issues emerging with counting too few or many owners with our three-person definition. This does not resolve the possibility of confusing employees with owners, to which we return shortly. Among the largest ownership teams for entering firms, we do not find immigrants being substantially different in terms of order listed. For four or more person teams, the immigrant shares are 20.3 percent, 14.4 percent, 13.8 percent, and 14.2 percent for the first through fourth positions (max reported) of listed owners. These structures again do not suggest very substantial issues likely to emerge with a three-person focus compared to using two- or four-person cut-offs.[7]

As a final step, the SBO contains some employment information that we can compare against the owner counts. As advanced warning, however, two significant issues exist in the analysis to follow. First, how each firm chooses to count owners toward employment is unclear. Second, the public-use SBO files intentionally introduce noise into the employment data to protect the identity of firms. Thus, while we believe the following analysis is quite informative for whether our definition creates a bias for immigrant versus native businesses, we need to be cautious about the exact figures reported. Our approach is simply to define an indicator variable for cases where we know we would have added an extra employee to our owner/entrepreneur definition because the employment count exceeds the owner count and the owner count was less than three owners; this is not comprehensive for all possible errors in our definition, but it is the most worrisome. Among the 2007 SBO entrants, this condition is met in 39.4 percent of cases. This number seems high to us, but we also do not know really how to evaluate it in light of the

7. In our regression sample, our mean immigrant entrepreneur share is 22.6 percent, with a 2005/1995 growth ratio of 1.31 (column [2] of table 5.3 that is discussed below). We find comparable means and very similar growth ratios when examining firms that start with one entrepreneur (23.6 percent, 1.32) or two entrepreneurs (21.4 percent, 1.29).

noisy data, and so forth. What we do take great comfort in, however, is that this fraction is 40.3 percent for immigrant-owned businesses and 39.1 percent for native-owned businesses, with the differences not statistically significant (*t*-stat = 0.74). This suggests that we while we may modestly mismeasure levels (e.g., identifying too many entrepreneurs), we are unlikely to have a bias by immigration status in this regard.

Overall, we take comfort in the SBO comparison. Any single rule like our three-person definition will face liabilities, and these tabulations appear to say our calibration is reasonably balanced. We also note one feature that helps isolate our technique from potential biases where we may identify an employee as an entrepreneur. Empirical studies of the hiring patterns by immigrant owners emphasize the strong degree to which they hire from their own ethnic group (e.g., Garca-Pérez 2012; Andersson, Burgess, and Lane 2012; Andersson et al. 2014; Åslund, Hensvik, and Skans 2014). Thus, our "false positive" for an immigrant-owned business is most likely to be an immigrant, and vice versa for a native-owned business. As an extreme example, our definition would be fully robust to unpaid owners or the inclusion of too many wage earners if the immigrant status of the employees exactly mirrored that of the true owners. This extreme condition does not hold, of course, but the quite high rates of concentration among small employers documented in Andersson et al. (2014) and similar studies are comforting for our design.

A second point of comparison comes from individuals to whom we can link LEHD records to their responses to the 2000 Census. The long form collects whether or not an individual declares themselves to be primarily self-employed in an incorporated firm, primarily self-employed in an unincorporated firm, an employee in a private-for-profit firm, or other categories. Looking at 2000 for individuals working in an SEIN created since 1995, we find that we label as an entrepreneur 66 percent of those declaring themselves to be self-employed in an incorporated firm. By contrast, we only label as an entrepreneur 29 percent of those declaring themselves to be self-employed in an unincorporated firm. Thus, our definition clearly tilts toward incorporated firms and those oriented toward employment and possibly growth, capturing a large portion of this group. The larger deviation, which is not surprising, is that about two-thirds of the overall set of people we declare to be entrepreneurs are listing themselves as an employee in a private-for-profit firm. Specifically, the composition of our entrepreneurial pool breaks down as 68 percent saying they are employed in private-for-profit firms, 28 percent self employed, and a small residual in other categories. In many of these new firms, no one is declaring themselves to be self-employed, which is a limitation of this approach to defining entrepreneurship. We thus find it difficult to benchmark this form of the metric compared to the self-employment breakdown.

5.3.4   Measuring Firm Dynamics

Our analyses consider the survival and growth of new businesses, which we measure exclusively through the LBD. This choice, which was not obvious to us at the start of this project, is important. By measuring outcomes through the LBD, we capture the full employment and payroll growth that the firm experiences (domestically in the United States). Alternative metrics based upon SEINs alone can miss substantial firm adjustments that occur within LEHD states when they open up new SEIN codes. Moreover, by definition, LEHD-based definitions of growth are subject to the state coverage limitations of the database. Thus, in combination with the discussion above, our strategy can be summarized as follows: (a) find single-unit firm entrants from 1992 to 2008 in the LBD that are in a covered LEHD state, (b) collect the initial employment records that are contained in the LEHD to describe the immigrant-native composition of initial employees and founders, and (c) return to the LBD for subsequent growth outcomes. This strategy allows us to retain all entrants and consistently measure growth; fortuitously, it also lets us use the LBD outcome data to 2011 for growth dynamics, even though our version of the LEHD ends in 2008. The only potential biases will relate to the specific set of states that we observe and how they compare nationally. On the other hand, this strategy would not necessarily be optimal in cases where one wanted to observe the employment composition of firm growth (e.g., the year-on-year subsequent hiring of immigrants or natives).

Our platform describes immigrant entrepreneurship in general and across subpopulations. We split the sample by low- and high-wage firms using the median quarterly earnings during the year of entry. We also define firms as high-tech if the majority of their employment is in a three-digit Standard Industrial Classification (SIC) code listed as a high-tech industry in Hecker (1999). Separate characterizations are also provided by one-digit SIC codes. Given the specific academic interest and policy focus on VC-backed firms, we identify entrants that receive VC funding by 2005, as recorded in the Venture Xpert database, using name- and geographical location-matching algorithms (Kerr, Nanda, and Rhodes-Kropf 2014). As we do not have matches beyond this point, we only study VC-backed entrepreneurship rates through 2005. Finally, we provide some tentative notes about whether firms go public by 2005. This information is collected by observing whether the new firm is later present in the Census Bureau's Compustat Bridge File, which was last updated for the 2005 public firm cohort.

5.3.5   Additional Discussion

A central goal of this project is the compilation of information and best practices necessary for using the LEHD for studies of immigrant entrepreneurship. To this end, a detailed data appendix provides specific instructions and commentary for researchers regarding the LEHD, with a special

focus on the firm-level dimensions and the longitudinal aspect of the LEHD data. This appendix information extends beyond the present study to also document issues noted in the Kerr, Kerr, and Lincoln (2015b) study of large firms. Additional restricted-access materials are available upon request. The appendix also provides thoughts about other data sets that could improve upon the entrepreneurship definitions developed here.

5.4   Entrepreneurship Trends

5.4.1   LEHD Statistics

Table 5.2A presents our core trends using all LEHD workers and a three-year definition for entrepreneurship. For these initial tabulations that regard all workers, we do not impose the single-unit firm restrictions or similar, keeping a very broad set of data. Column (2) considers immigrant participation rates in new firms relative to the total immigrant workforce in the LEHD. Approximately 6.0 percent of immigrants in the LEHD are working in new firms born over the prior three years, with some evidence for a decline in the rate over time. Appendix table 5A.2A provides the observation counts that underlie these estimates, which cover 3.2 and 4.6 million immigrants in 1995 and 2008, respectively. Throughout this chapter, observation counts are approximate and rounded per Census Bureau disclosure requirements. This appendix table also shows that immigrants constitute 19.3 percent of the LEHD sample from 1995 to 2008, growing from 16.4 percent to 21.2 percent, and the native rate of employment in new firms averages 4.6 percent (versus 6.0 percent for immigrants). The native rate is similarly declining after 2005.[8]

Column (3) documents that 2.2 percent of immigrants are among the top earners in a new business and thus declared to be an entrepreneur by our definition. Parallel to Fairlie and Lofstrom (2014) and in contrast to column (2), this share grows substantially to 2005 before declining some. Features of the data initialization process for identifying entrepreneurship are more difficult for 1995 and 1996 for some states, and we are required to use some minor extrapolation for these initial years for column (3) in tables 5.2A and 5.2B. These initialization challenges impact entrepreneurial rate calculations mostly, with upcoming share-based calculations in columns (4)–(10) being substantially less sensitive and unadjusted. Similarly, we report the trend statistics through 2008, but we hesitate to make too much of the declines observed after 2005. Through 2005, the entry rates are overall stable, and we believe some, if not all, of the decline after 2005 can be traced to declines in

8. While the eleven states are present in the LEHD by 1992, the statistics begin in 1995 to allow full initiation of all of our data and definitions. For example, while we can identify from the LBD the full set of young firms in 1992 (i.e., those born in 1989–1992), we do not know the immigration status of all of their top earners in the first year of the firm's life.

Table 5.2A   Trends in immigrant entrepreneurship and participation in new firms

Year (1)	Immigrant rate of employment in new firms (2)	Immigrant entrepreneur rate, defined as top three initial earners in business (3)	Immigrant share of employment in new firms (4)	Immigrant entrepreneur share, defined as top three initial earners in business (5)	Column (5) restricted to top quartile of entrepreneurial income in start year (6)	Share of entering LEHD SEINs with one or more immigrant (7)	Share of entering LBD firms with one or more immigrant (8)	Immigrant share of workers in entering LBD firms (9)	Column (9) weighted by employment (10)
1995	6.3	1.6	20.6	16.7	15.6	29.8	31.1	19.3	18.8
1996	6.4	2.0	21.0	19.3	15.5	30.9	33.1	20.3	19.4
1997	6.2	2.1	21.5	21.0	16.3	31.7	33.0	20.5	19.8
1998	6.1	2.1	21.9	22.1	16.9	32.4	33.8	20.8	20.0
1999	6.2	2.1	22.0	22.6	16.9	33.5	34.3	21.3	20.2
2000	6.1	2.1	22.7	23.2	18.5	34.2	35.9	22.1	20.9
2001	6.0	2.1	23.5	24.3	18.8	35.6	36.6	23.6	22.6
2002	5.9	2.2	24.7	25.6	20.0	36.7	38.5	25.1	23.4
2003	6.1	2.3	25.2	26.6	19.4	36.0	38.6	25.7	24.0
2004	6.3	2.4	25.7	27.0	20.3	35.8	36.8	24.9	24.1
2005	6.3	2.5	25.7	27.1	20.7	35.9	37.4	25.3	24.0
2006	6.0	2.4	25.6	27.2	19.8	35.9	37.8	26.3	24.4
2007	5.5	2.3	25.4	26.9	19.7	34.2	36.0	25.2	24.4
2008	4.9	2.0	25.6	27.1	20.1	35.6	37.0	26.2	24.8
Mean	6.0	2.2	23.7	24.1	18.5	34.2	35.7	23.3	22.2
Ratio 08/95	0.78	1.24	1.24	1.62	1.29	1.19	1.19	1.36	1.32
Ratio 05/97	1.02	1.22	1.20	1.29	1.27	1.13	1.13	1.23	1.21

Notes: Table provides broad trends related to employment in new firms by natives and immigrants in the LEHD and LBD. The sample includes eleven states present in the LEHD by 1992: CA, CO, FL, ID, IL, LA, MD, NC, OR, WA, and WI. Columns (2) and (3) consider immigrant participation rates in new firms relative to the total immigrant workforce in the LEHD. Columns (4)–(10) consider immigrant shares of activities relative to natives. New firms are defined through the LBD as described in the text and retain their entering status for the first three years of the firm's life. Caution should be exercised with column (2) and (3)'s trends from 2006 onward due to declines in match rates for the Business Registry Bridge between the LEHD and LBD late in the sample. Caution should also be used for column (3)'s entry rates in 1995 and 1996 due to ongoing initialization that required minor extrapolation. Share-based calculations in columns (4)–(10) are substantially less sensitive to these issues. The appendix provides complementary statistics: (a) considering shorter series for twenty-eight states present in the LEHD by 2000, (b) adjusting the definition of new firms to apply to the first year of business entry only, and (c) separating one-digit SIC industries.

match rates of new firms in the Business Registry Bridge between the LEHD and LBD. That said, to the extent that the trends are real, they would match the broad secular decline in employment in young US firms documented by Decker et al. (2014). Most of our focus is on the share of entrepreneurs who are immigrants, which is not materially influenced by these issues.

Columns (4)–(10) consider immigrant shares of activities relative to natives. Columns (4) and (5) repeat the previous two analyses on a relative basis. We estimate immigrants account for about 24 percent of entrepreneurs and the new employees of firms in our sample. The immigrant share of new entrepreneurs rises dramatically in our sample from 16.7 percent in 1995 to 27.1 percent in 2008, while the trend for immigrant shares of initial employees is more modest. Figure 5.2 graphs these trends. In column (6), we isolate the top quartile of the initial earnings distribution of start-ups (measured as quarterly averages). Immigrants tend to create firms with lower initial earnings, and the upward trend in immigrant shares for this top quartile is a bit weaker. Some of this pattern resurfaces below when looking at payroll growth regressions.

A number of studies report the share of firms with at least one immigrant founder. This often appears motivated by a desire to have the highest share possible for advocacy pieces, but it may also stem from a genuine
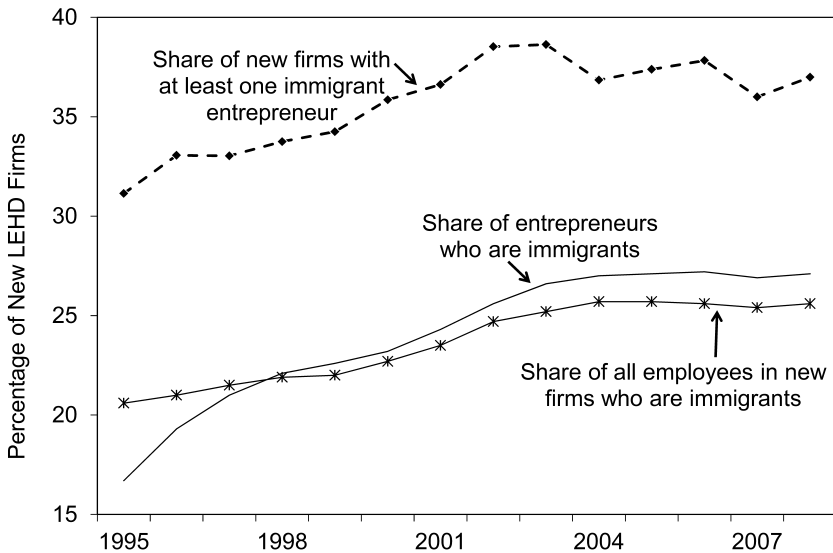
Fig. 5.2   LEHD immigrant entrepreneurship trends

Notes: See table 5.2A. Sample includes eleven states present in the LEHD by 1992: CA, CO, FL, ID, IL, LA, MD, NC, OR, WA, and WI. New firms are defined through the LBD and retain their entering status for the first three years of the firm's life. Entrepreneurs are defined as top three initial earners in business.

desire to capture the number of businesses touched in some way by immigrant entrepreneurship. By contrast, our baseline estimates are implicitly allowing fraction immigrant entrepreneurship in firms with several starting employees and weighting larger initial teams more (up to three employees). Columns (7) and (8) show that about 35 percent of entering SEINs or LBD firms have an initial immigrant employee, with column (7) also implicitly showing that our LEHD-LBD match is not introducing a bias. Columns (9) and (10) show patterns defined over larger LBD samples. While the results vary somewhat depending on these definitions used, the main message from table 5.2A is that the overall time-series pattern of our findings remains largely unchanged—immigrants are accounting for roughly a quarter of entrepreneurs and their share is increasing with time.

Appendix table 5A.2A provides complementary statistics for the twenty-eight states present in the LEHD by 2000. The wider state panel allows us to assess the potential impact of focusing on just eleven states, including two of the nation's most immigrant-heavy states, California and Florida. Considering the overlapping 2002–2008 period, our longer panel has an average immigrant worker share of 20.6 percent compared to 17.3 percent in the wider sample; likewise, the immigrant share of entrepreneurship is 25.4 percent versus 21.6 percent. Thus, our "levels" statistics are on the order of 3 percent higher. On the other hand, rates of immigrant and native entry are extremely close (5.9 percent vs. 5.8 percent, 4.5 percent vs. 4.4 percent), and all of the 2002–2008 trends are very close to each or even stronger in the wider sample. We thus conclude that our longer state sample may overstate national levels slightly, but is otherwise quite representative. This is due in part to the larger average state sizes in the longer panel (despite the addition of Texas in the larger set of twenty-eight states), with the eleven baseline states constituting 57 percent of the employment in the twenty-eight states from 2002 to 2008.

Appendix table 5A.2B shows that the results are robust to defining new firm employment through the first year of business entry only. This narrower definition essentially cuts the rate of firm entry by two-thirds. The one-year employment rate in new firms for immigrants is now 2.0 percent, compared to 1.5 percent for natives, a sizable differential remaining. The immigrant share of employment in new firms grows from 17.6 percent in 1992 to 24.9 percent in 2008, parallel to our baseline results in table 5.2A.

Appendix tables 5A.3A and 5A.3B report results for one-digit industries. Rates of immigrant entry are highest in mining and construction (SIC1), wholesale and retail trade (SIC5), and services (SIC7). Splitting industries at the three-digit level, entry rates tend to be higher for low-tech sectors, but this pattern is inverted around 2000 during the high-tech boom. A similar pattern is evident for VC-backed entry. Some of these patterns reflect inherent differences in entry rates across sectors and over time, common to immigrants and natives. In terms of share of initial employment, immigrants have

higher relative representation in manufacturing (SIC2–3), wholesale and retail trade (SIC5), and services (SIC7).

Table 5.2B repeats table 5.2A for the part of our sample of firms that are backed by VC investors. About 0.11 percent of immigrants are starting top earners in VC-backed firms from 1995 to 2005. This share is naturally substantially less than the 2.2 percent in table 5.2A due to the relatively small number of ventures receiving VC investment. Reflecting the VC bubble that peaked in 2000, this rate is hump-shaped over time with a peak in 2000. Immigrants constitute about 28 percent of VC-backed founders, with this fraction increasing over time. The substantial majority of entering VC-backed firms have at least one initial immigrant employee, with the more similar results regarding overall fractions of founders coming from the fact that VC-backed firms tend to have larger counts of initial employees and larger founding teams. On the whole, immigrant entrepreneurship is somewhat stronger for VC-backed firms than generally, with 30 percent of the former being immigrants compared to 27 percent overall in 2005.

Our data platform allows us to compare initial immigrant and native employees by (a) their LEHD characteristics for the full sample, and (b) the Census long-form responses for the matched sample. Appendix tables 5A.4 and 5A.5 provide these results that we quickly summarize here. On average, immigrant employees in new firms tend to be slightly older and more likely to be male, with lower average LEHD quarterly earnings before, during, and after the founding of the firm. By contrast, the average quarterly earnings for immigrants before, during, and after the founding of VC-backed firms tends to be higher than their native peers. Looking at respondents matched to the 2000 Census, immigrants employed in young firms are more likely to be older, male, married, and have more children. While less likely to own a home, immigrants are more likely to own a higher-priced home and rent more expensive properties. This is partially connected to immigrants being more likely to live in high-priced gateway cities. The average year of arrival for our 2000 cross section is 1984, so that the average tenure in the United States in 2000 is sixteen years. These statistics are reasonable and comforting for our match, although we focus most of our analytical attention elsewhere. We return later to the year of arrival when considering differences between adult arrivals and those migrating as children.

Fairlie (2013) documents from the 2007 SBO that immigrant-owned businesses represent 13.2 percent of all businesses in the United States, with $434,000 in average annual sales (compared to non-immigrant-owned firm sales of $609,000). Only 28 percent of immigrant-owned businesses in the SBO hire outside employees, while the share is even smaller (26 percent) among the non-immigrant-owned businesses. For those that do hire employees, the average number of employees is eight in the immigrant-owned businesses and twelve in the native-owned businesses. Among our sample, the average initial employment for firms founded by immigrants exclusively

Table 5.2B Table 5.2A for firms backed by venture capital

Year (1)	Immigrant rate of employment in new firms (2)	Immigrant entrepreneur rate, defined as top three initial earners in business (3)	Immigrant share of employment in new firms (4)	Immigrant entrepreneur share, defined as top three initial earners in business (5)	Column (5) restricted to top quartile of entrepreneurial income in start year (6)	Share of entering LEHD SEINs with one or more immigrant (7)	Share of entering LBD firms with one or more immigrant (8)	Immigrant share of workers in entering LBD firms (9)	Column (9) weighted by employment (10)
1995	1.1	0.06	28.3	24.7	20.1	50.8	61.6	23.6	18.6
1996	1.4	0.08	28.4	24.7	25.1	49.6	73.3	25.5	22.7
1997	1.8	0.09	24.3	25.6	25.1	53.7	64.4	23.6	20.8
1998	1.6	0.12	24.3	27.0	24.0	55.5	74.1	28.9	24.4
1999	2.0	0.14	24.2	27.1	27.4	57.5	76.1	28.4	26.1
2000	2.2	0.17	23.5	27.8	26.4	62.0	87.9	31.7	27.4
2001	2.1	0.16	26.0	30.5	38.5	57.2	76.1	33.6	31.0
2002	1.7	0.13	26.1	32.9	31.5	59.1	83.0	37.7	35.7
2003	1.4	0.11	27.2	31.2	33.1	54.8	77.7	35.4	36.0
2004	1.3	0.10	24.6	30.4	31.3	59.9	79.3	35.7	33.9
2005	1.9	0.10	28.5	30.0	29.1	57.9	79.4	34.9	33.3
Mean	1.7	0.11	25.9	28.4	28.3	56.2	75.7	30.8	28.2
Ratio 05/95	1.73	1.71	1.01	1.21	1.45	1.14	1.29	1.48	1.79
Ratio 05/97	1.06	1.08	1.17	1.17	1.16	1.08	1.23	1.48	1.60

Notes: See table 5.2A Companies backed by venture capital investors are identified through firm names and locations as described in the text. Venture capital matching is available through 2005. Inclusion in the venture capital group occurs if venture capital financing is ever received, rather than being specific to the first years of the firm.

is 4.4 workers, compared to 7.0 workers for firms launched exclusively by natives. When both types of founders are present, the average is 16.9 workers. Thus, in general, our data appears to match the broad levels and patterns of the employer firms in the SBO sample, as well as differences in the typical sizes between immigrants and natives. Our overall numbers are naturally lower given the focus on the initial year of the firm (versus a cross section of employment patterns in existing firms).
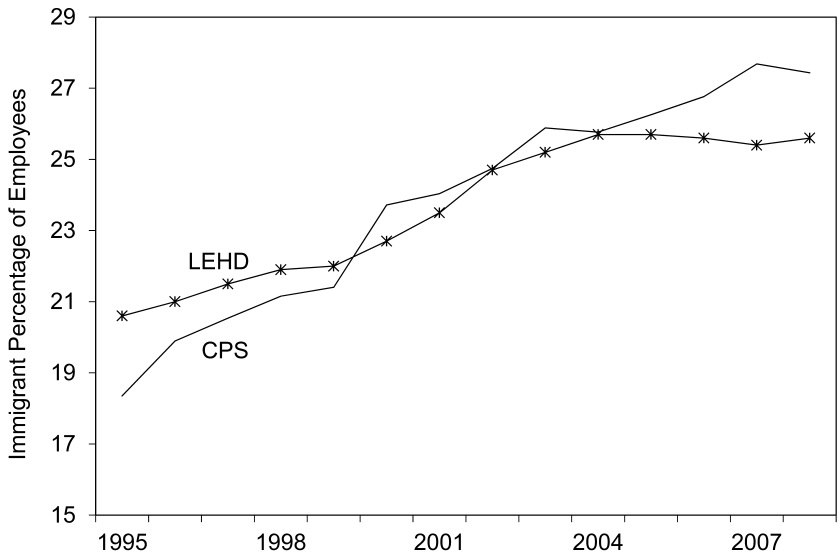
5.4.2   CPS Comparison

To check our LEHD-based approach against publicly available survey data, we derive entrepreneurship trends using the Current Population Survey (CPS) Merged Outgoing Rotation Group (MORG) data from the NBER.[9] The data include details about the respondent's place of birth starting in 1994, and also report the class of worker where one of the categories is "self-employed in an incorporated business." We prepare a CPS sample that matches the traits of our LEHD work, starting with individuals age twenty-five to fifty who live in one of our core eleven states and work in the private sector. We further limit the data to those persons who are in the labor force and have a known level of education and potential labor market experience of at least one year. To stay consistent with the LEHD, we include as immigrants all those who are born outside of the United States.

Figure 5.3A first compares the immigrant workforce shares evident in the two data sources. The levels are very comparable, and the trends quite similar, with the CPS trend being modestly sharper. Figure 5.3B next compares the immigrant share of the entrepreneurial groups. The CPS trend includes as new entrepreneurs those who are newly self-employed in an incorporated business. The levels are more different here, with the LEHD being consistently at least 3 percent–4 percent higher. Perhaps more important, while both are upwardly sloping, the timings are different. The LEHD shows stronger growth during the 1990s, while the CPS picks up more in the middle to late first decade of the twenty-first century. We do not have a strong hypothesis regarding the source of these differences, although some of it may connect to the CPS redesign in 2002–2003. Unreported tabulations include unincorporated self-employed workers into the CPS entrepreneurial definitions, finding that the resulting trend sits in between the two series shown in figure 5.3B, with the augmented CPS series retaining its trend differences from the LEHD.

Figure 5.3C finally compares various metrics regarding rates of entrepreneurship for immigrants. Relative to the immigrant entrepreneurial shares shown in figure 5.3B, entrepreneurial rate calculations are substantially more sensitive to definitional decisions that can have large impact on their levels (regardless of data source). From the CPS, we provide at the top of

---

9. We thank Ethan Lewis for his guidance on this work. (Data available at: http://www.nber .org/data/morg.html.)

Fig. 5.3A   LEHD-CPS comparison for immigrant worker shares (workers in the private sector in eleven focal LEHD states)

Notes: Samples from both data sources include eleven states present in the LEHD by 1992: CA, CO, FL, ID, IL, LA, MD, NC, OR, WA, and WI. Included individuals are age twenty-five to fifty, employed in the private sector, and meet certain educational and work history restrictions. An immigrant is defined as a person born outside of the United States.
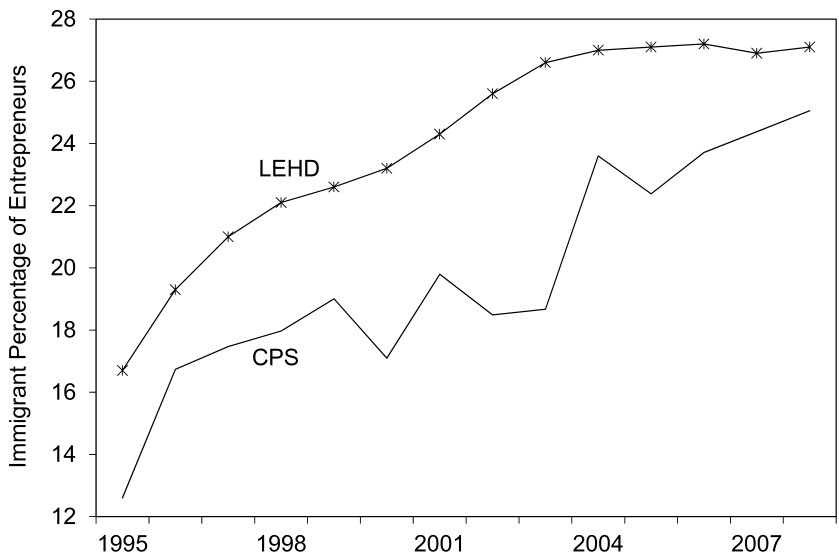
Fig. 5.3B   LEHD-CPS for immigrant entrepreneurial shares (entrepreneurship and self-employment in the private sector in eleven focal LEHD states)

Notes: See figure 5.3A. LEHD entrepreneurs are defined as top three initial earners in business and retain this status for the first three years after the firm's start. CPS entrepreneurs are defined as those entering incorporated self-employment.
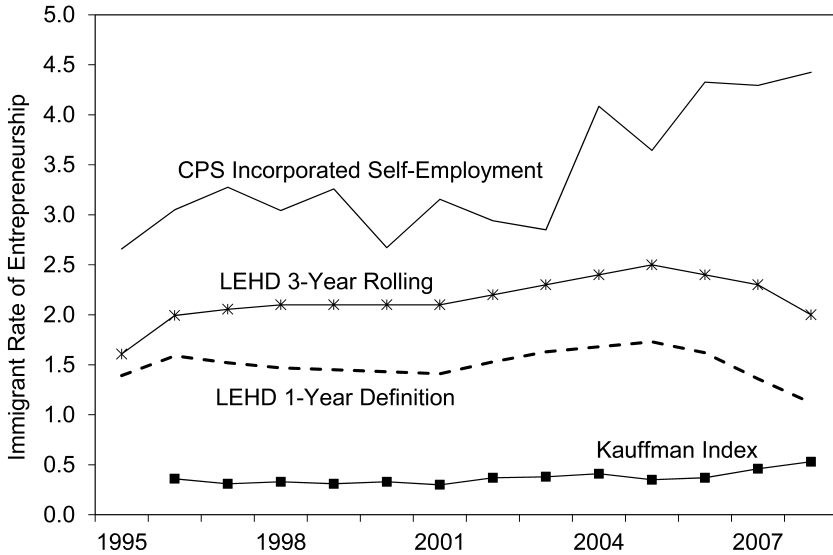
Fig. 5.3C   LEHD-CPS for immigrant entrepreneurial rates (comparison of immigrant entrepreneurial measures to immigrant population)

Notes: CPS measures include the overall incorporated self-employment rate for immigrants in the sample, which retains self-employed owners who have held their business for many years, and the Kauffman Foundation's Index that is derived from the CPS through new entry into self-employment. LEHD measures are provided with one- and three-year windows for comparison.

figure 5.3C a measure of the incorporated self-employment rate for immigrants in the sample. This includes self-employed owners who have held their business for many years, and thus provides a higher estimate of about 3.5 percent for the years; this share would exceed 10 percent if including unincorporated self-employment. At the bottom of figure 5.3C is the Kauffman Foundation Index that is derived from the CPS through entry into self-employment (Fairlie et al. 2015); due to its focus on entrepreneurial transitions, the rate is much smaller at about 0.4 percent. For the LEHD we show our core measure, where we use a three-year window for counting entrepreneurs and the one-year version that is most comparable to the Kauffman Foundation Index. It is clear that our metrics fall in between the extremes of CPS-based approaches. We tend to see comparable stability, and both data sources speak to a growing rate for immigrants compared to natives in terms of entrepreneurial transitions (which is mostly implied by figure 5.3B).[10]

10. The one-year rolling definition of entrepreneurship in the LEHD provides an entry rate that is about two-thirds of the three-year basis in figure 5.3C. This limited gap is due to the high rate of business failure in the first three years of operation. By contrast, employment in new ventures shows a greater difference, as described above for appendix table 5.2B. This is because employment counts capture the growth and scaling of surviving ventures through their first

5.5   Analysis of Firm Survival and Growth

We next consider differences in the performance of firms founded by immigrants versus natives. Table 5.3 first provides descriptive statistics for the sample of firms used for analytical work comparable to tables 5.2A and 5.2B. The analytical sample includes firms founded 1992–2005 in a Primary Metropolitan Statistical Area (PMSA) within a state present in the LEHD since 1992. Relative to tables 5.2A and 5.2B, several data preparation steps are undertaken to exclude entrants that are multiunit LBD entering firms and entrants lacking complete information for the considered LBD and LEHD metrics (e.g., reported payroll). The metrics focus on the immigrant-native composition of the top three initial earners. The sample ends with 2005 entrants to allow observation of LBD outcomes until 2011 and to circumvent any issues with the LEHD-LBD match (Business Registry Bridge) in later years.

Table 5.4A shows a basic tabulation of the data over a three-year growth horizon. We include in this analysis all entering cohorts of firms during the 1992–2005 period, with outcomes measured after three years for each entrant (e.g., 2004 employment for a 2001 entrant). In panel A, each row represents a separate starting size category in terms of employment. We then tabulate the share of entrants for each starting size category that grow to the level indicated by column headers by the third year, with rows summing to 100 percent. Thus, column (2) shows that over one-third of firms close across this time span, while column (7) shows that very few firms reach or exceed 100 employees, especially when starting from the smallest size category. The cells in bold represent the least moment from initial employment levels, which is the most likely outcome other than business closure.[11]

Panel B provides for each cell the average initial immigrant entrepreneurship share for the firms in that group. Entrepreneur definitions use the top three initial earners, independent of whether these individuals remain top earners in the firm across the three years. The Total rows and columns provide the weighted average immigrant entrepreneurship rates for their groups. The shares in column (8) decline across starting levels reflective of the lower sample average initial employment for immigrant-started businesses noted above. Panel C similarly provides the average initial immigrant employment share for grouped firms independent of initial earnings.

The intriguing initial pattern in the data points to a greater volatility of immigrant entrepreneur outcomes. The immigrant shares are frequently lowest among the bolded cells that represent static employment levels. In all cases, immigrants are more represented among closed firms in column (2)

---

three years, in addition to business failure effects. As our identification of entrepreneurs is fixed from the first year of each venture and with a maximum of three founders, this latter effect is not present and the differences based upon windows is smaller in figure 5.3C.

11. The majority of closed businesses are failed companies, but closures also include acquired companies should the LBD identifiers change, some of which may be quite successful outcomes.

Table 5.3   Trends for samples used in analytical work

Year (1)	Full sample			Firms backed by venture capital		
	Immigrant entrepreneur share, defined as top three initial earners in business (2)	Column (2) with employment weights (3)	Share of entering firms with one or more immigrant entrepreneur (4)	Immigrant entrepreneur share, defined as top three initial earners in business (5)	Column (2) with employment weights (6)	Share of entering firms with one or more immigrant entrepreneur (7)
1995	19.2	18.1	23.7	22.3	15.2	35.6
1996	20.2	18.4	25.1	21.9	17.4	42.6
1997	20.5	19.3	25.5	25.2	23.3	39.4
1998	20.7	19.2	25.9	28.1	17.0	44.0
1999	21.2	19.4	26.7	28.0	25.1	48.5
2000	22.2	20.6	28.1	28.7	21.6	49.4
2001	23.6	21.9	29.4	30.1	24.5	53.8
2002	25.0	22.3	30.8	34.4	26.8	58.0
2003	25.6	23.3	31.3	30.9	25.1	52.1
2004	24.9	23.5	30.3	32.0	22.7	57.5
2005	25.2	23.2	30.6	31.2	27.0	49.5
Mean	22.6	20.8	27.9	28.4	22.3	48.2
Ratio 05/95	1.31	1.28	1.29	1.40	1.78	1.39
Ratio 05/97	1.23	1.20	1.20	1.24	1.16	1.26

Notes: Table provides descriptive statistics for the sample of firms used for analytical work. The sample includes firms founded 1992–2005 in a PMSA within a state present in the LEHD since 1992. Relative to tables 5.2A and 5.2B, several data preparation steps are undertaken to exclude entrants that are multiunit LBD entering firms and entrants lacking complete information for considered LBD and LEHD metrics (e.g., reported payroll). Metrics focus on the immigrant-native composition of the top three initial earners. The sample ends with 2005 entrants to allow observation of LBD outcomes to 2011 and to circumvent issues with the LEHD-LBD match in later years. Employment weights are capped at fifty initial employees.

Table 5.4A **Growth tabulations and immigrant entrepreneurship share, three-year horizon**

Row headers group by initial employment: (1)	Employment after three years:						
	Closed (2)	1–4 empl. (3)	5–9 empl. (4)	10–19 empl. (5)	20–99 empl. (6)	100+ empl. (7)	Total (8)
A. Distribution of outcomes by initial firm size, full sample (rows sum to 100 percent)							
1–4 employees	0.392	**0.448**	0.110	0.034	0.014	0.001	
5–9 employees	0.356	0.167	**0.281**	0.144	0.049	0.003	
10+ employees	0.385	0.040	0.087	**0.209**	0.246	0.033	
B. Starting immigrant entrepreneurship share in new firms by outcome distribution							
1–4 employees	0.226	**0.220**	0.205	0.200	0.205	0.212	0.220
5–9 employees	0.208	0.206	**0.180**	0.182	0.213	0.187	0.196
10+ employees	0.200	0.207	0.202	**0.179**	0.181	0.166	0.190
Total	0.220	0.219	0.196	0.187	0.191	0.173	0.212
C. Starting immigrant employment share in new firms by outcome distribution							
1–4 employees	0.230	**0.224**	0.209	0.205	0.213	0.240	0.224
5–9 employees	0.213	0.209	**0.183**	0.187	0.221	0.218	0.200
10+ employees	0.215	0.217	0.203	**0.182**	0.197	0.232	0.203
Total	0.226	0.222	0.199	0.191	0.204	0.232	0.217

Notes: Table describes transition/growth properties over three-year horizon for entering cohorts 1992–2005, with outcomes measured after three years for each entrant (e.g., 2004 for a 2001 entrant). Panel A tabulates the share of entrants for each starting size category that grow to the level indicated by column headers, with rows summing to 100 percent. Panel B provides for each cell the average initial immigrant entrepreneurship share for grouped firms. Entrepreneur definitions use top three initial earners, independent of whether these individuals remain with the firm or as a top earner in the firm. Panel C provides for each cell the average initial immigrant employment share for grouped firms.

compared to the bolded cells. Moreover, among the firms that begin with 5–9 or 10 or more employees, immigrant shares are lowest in the bolded cells compared to any other movement. In the smallest category, which panel A shows is the most stagnant of the initial size categories, immigrant shares among those firms remaining at 1–4 employees closely mirrors the overall share. Table 5.4B similarly considers a six-year growth horizon for cohorts, which uses the LBD data up to 2011 for our 2005 entrants. Over half of entrants fail by year six, which reflects typical start-up life expectancies. The patterns are mostly repeated here, especially in panel C. In panel B, there is less evidence of upside growth outcomes.

While intriguing, these tabulations do not account for general differences in when immigrant versus native firms are founded or their other measurable attributes. To address these issues, table 5.5A considers regressions of the three-year outcomes that take the form,

$$Y_{f,t+3} = \eta_t + \beta \text{ImmigrantEntrepreneurShare}_{f,t} + \gamma X_{f,t} + \varepsilon_{f,t},$$

where f and t index firms and entering cohorts. We include a vector of cohort fixed effects η_t and control for the initial attributes of the firms ($X_{f,t}$) in terms of their starting log employment and log payroll. Regressions are unweighted and report robust standard errors. Panel A presents the summary statistics for outcome variables. Panel B provides the baseline regression, with the last row giving the relative effect of increasing the immigrant entrepreneur share from zero to one compared to the sample mean.

Column (1) considers firm survival until the third year. On average, 64 percent of firms survive this long, with immigrant-founded firms being 0.3 percent less likely to survive compared to similar firms with only native founders. Columns (2)–(7) consider traits of firms conditional on surviving to their third year. We first look at employment growth, measuring changes relative to the firm's average in two periods following Davis, Haltiwanger, and Shuh (1996): $[Y_{f,t+3} - Y_{f,t}] / [(Y_{f,t+3} + Y_{f,t}) / 2]$. This measure is bounded by $[-2, 2]$, prevents outliers, and symmetrically treats positive and negative shifts. Conditional on survival, firms founded by immigrants show greater growth. Columns (3) and (4), by contrast, show no difference in terms of growth of payroll or establishment counts. The lower payroll growth may indicate lower wage growth, additional jobs being lower wage, or that partial employment is in greater use. Column (5) alternatively models employment growth through indicator variables for the firm achieving more workers than its initial level, while columns (6) and (7) are similarly defined for the firm reaching 100 workers or being among the top 10 percent of firms in terms of employment in its specific industry. These approaches confirm the employment growth observed in column (2). The final regression shows immigrant-started firms are much more likely to receive VC financing.

Panel C takes a more stringent approach that controls for cohort-PMSA-industry fixed effects, with industry being defined at the two-digit

Table 5.4B Table 5.4A with a six-year horizon

Row headers group by initial employment: (1)	Employment after six years:						
	Closed (2)	1–4 empl. (3)	5–9 empl. (4)	10–19 empl. (5)	20–99 empl. (6)	100+ empl. (7)	Total (8)
A. Distribution of outcomes by initial firm size, full sample (rows sum to 100 percent)							
1–4 employees	0.571	**0.299**	0.082	0.032	0.015	0.001	
5–9 employees	0.536	0.130	**0.180**	0.105	0.046	0.003	
10+ employees	0.566	0.037	0.065	**0.136**	0.171	0.024	
B. Starting immigrant entrepreneurship share in new firms by outcome distribution							
1–4 employees	0.226	**0.215**	0.203	0.198	0.201	0.221	0.220
5–9 employees	0.208	0.192	**0.178**	0.174	0.203	0.178	0.196
10+ employees	0.198	0.199	0.206	**0.175**	0.173	0.155	0.190
Total	0.220	0.213	0.196	0.183	0.186	0.170	0.212
C. Starting immigrant employment share in new firms by outcome distribution							
1–4 employees	0.230	**0.219**	0.206	0.203	0.209	0.241	0.224
5–9 employees	0.212	0.195	**0.179**	0.180	0.209	0.220	0.200
10+ employees	0.214	0.205	0.204	**0.177**	0.189	0.215	0.203
Total	0.225	0.216	0.198	0.188	0.198	0.221	0.217

Notes: See table 5.4A. Outcome distributions use a six-year horizon for each entrant (e.g., 2007 for a 2001 entrant).

Table 5.5A **Regressions of three-year outcomes using LBD firms**

	Fraction alive in third year (1)	Employment growth relative to firm's average (2)	Payroll growth relative to firm's average (3)	Establishment count growth relative to firm's average (4)	(0,1) third-year employment exceeds initial level (5)	(0,1) third-year employment exceeds 100 workers (6)	(0,1) third-year employment is in top decile of industry (7)	(0,1) received venture capital support (8)
					Traits of surviving firms in the third year			
A. Summary statistics								
Mean	0.6418	0.2726	0.2449	0.0033	0.5133	0.0086	0.0848	0.0027
SD	0.4866	0.6823	0.7216	0.0567	0.4998	0.0921	0.2787	0.0515
Observations	535,580	329,260	329,260	329,260	329,260	329,260	329,260	535,580
B. Estimation with cohort fixed effects and controls for initial traits								
Firm's immigrant entrepreneur share	−0.00335	0.03130	0.00037	−0.00025	0.01919	0.00159	0.00362	0.00232
	(0.00186)	(0.00301)	(0.00356)	(0.00022)	(0.00241)	(0.00036)	(0.00109)	(0.00020)
ß/mean of DV	−0.005	0.115	0.002	−0.076	0.037	0.185	0.043	0.872
C. Estimation with cohort-PMSA-industry fixed effects and controls for initial traits								
Firm's immigrant entrepreneur share	0.01851	0.00190	−0.02375	−0.00024	0.00656	0.00047	0.00298	0.00045
	(0.00247)	(0.00483)	(0.00569)	(0.00036)	(0.00384)	(0.00064)	(0.00184)	(0.00031)
ß/mean of DV	0.029	0.007	−0.097	−0.073	0.013	0.055	0.035	0.169

Notes: See tables 5.3 and 5.4A. Table quantifies how the outcomes of new firms vary by the share of the top three initial earners who are immigrants. The sample includes firms founded 1992–2005 in a PMSA within a state present in LEHD since 1991. Outcome variables are measured through 2011 using the LBD. Columns (2)–(4) measure growth by comparing the change during the period to the average of the start and end values for the firm. Regressions are unweighted, report robust standard errors, and control for the initial traits of starting log employees and log payroll of the venture. Observation counts are approximated per Census Bureau requirements.

level of the SIC classification. This approach removes any differences between immigrants and natives to found firms in certain cities or sectors, which can be important for outcomes, and instead compares immigrant outcomes to natives within these narrow cells. We do not view either approach as an inherently better way to describe the data, as differences in the choices of locations and industries are as relevant as differences in within-cell outcomes (e.g., due to different management practices). In some cases, this choice has material impacts, while in other cases the results are robust across the variation employed in the estimates.[12]

In panel C, we see that immigrant entrepreneurs are more likely to survive for three years compared to their closest native peers. The baseline employment growth is substantially diminished in column (2) compared to panel A, while some of the binary employment growth outcomes in columns (5)–(7) retain more strength. The most robust outcome is achieving the top decile of industry employment, which already has a degree of industry-level benchmarking built into it. Payroll growth is now significantly less than native peers, while establishment count growth is again flat.

Table 5.5B repeats table 5.5A for the six-year outcomes. The relationships in panel B are quite comparable to the three-year outcomes, with payroll and establishment count growth now more evident. Payroll growth is again noticeably smaller than employment growth. In general, there is greater statistical precision for the results with six-year outcomes, and the relative magnitudes are larger in size compared to the US average outcomes than earlier. The introduction of cohort-PMSA-industry fixed effects in panel C of table 5.5B has a similar effect to what it did for the three-year outcomes. Appendix tables 5A.6A and 5A.6B repeat these outcomes using simply the initial immigrant employee share as the explanatory variable, finding comparable results. Perhaps the key difference is that the employment growth outcomes are now more robust to the inclusion of the cohort-PMSA-industry fixed effects.

On the whole, the data paint some interesting differences, albeit tentative and noncausal, between firms founded by immigrants versus natives. When incorporating industry and city choice into the variation, immigrant founders have a greater volatility that somewhat mimics the up-or-out dynamics of young firm growth described in Haltiwanger, Jarmin, and Miranda (2013). They fail more frequently, but generate greater employment growth if they manage to stay in business. Over a six-year horizon they become more associated with higher payroll and establishment counts, but these are second

---

12. Our analysis uses the geocoding for the initial PMSA for the business. In a separate context, Kerr, Kerr, and Nanda (2015) describe the geographic information in much greater detail for the LEHD and the LEHD-Decennial match. There is capacity within the LEHD to observe movements over cities for all individuals; for those matched to the Decennial Census, there is further power to look very closely at the locations of residence versus business. These dimensions would be very interesting to consider in the immigrant context.

Table 5.5B Table 5.5A with a six-year horizon

	Fraction alive in sixth year (1)	Employment growth relative to firm's average (2)	Traits of surviving firms in the sixth year				
			Payroll growth relative to firm's average (3)	Establishment count growth relative to firm's average (4)	(0,1) sixth-year employment exceeds initial level (5)	(0,1) sixth-year employment exceeds 100 workers (6)	(0,1) sixth-year employment is in top decile of industry (7)
A. Summary statistics							
Mean	0.4350	0.3097	0.3643	0.0034	0.5362	0.0102	0.0842
SD	0.4958	0.7403	0.7984	0.0591	0.4987	0.1003	0.2776
Observations	535,580	233,000	233,000	233,000	233,000	233,000	233,000
B. Estimation with cohort fixed effects and controls for initial traits							
Firm's immigrant entrepreneur share	−0.01365	0.04678	0.00876	0.00047	0.02899	0.00233	0.00386
	(0.00188)	(0.00393)	(0.00468)	(0.00029)	(0.00290)	(0.00048)	(0.00134)
ß/US mean	−0.031	0.151	0.024	0.138	0.054	0.228	0.046
C. Estimation with cohort-PMSA-industry fixed effects and controls for initial traits							
Firm's immigrant entrepreneur share	0.01795	0.00760	−0.02769	−0.00032	0.01326	0.00157	0.00029
	(0.00243)	(0.00665)	(0.00794)	(0.00038)	(0.00489)	(0.00087)	(0.00245)
ß/US mean	0.041	0.025	−0.076	−0.094	0.025	0.154	0.003

Notes: See table 5.5A

order to employment outcomes. However, much of this action appears to come through the way in which immigrant entrepreneurs chose locations and industries. Conditional on these features, they are more likely to survive than their native peers and modestly more likely to experience employment growth, with payroll growth underperforming comparable firms founded by natives.

Unreported estimations consider whether the industry or geography controls are more important for the differences observed across panels in tables 5.5A and 5.5B. The intriguing answer is that geography plays the central role—especially one state. This can be expressed in two ways. First, where a reversal of coefficient direction occurs across panels, the same pattern usually occurs when just introducing state fixed effects, while this is not true when introducing industry fixed effects. Second, adjustments in sample composition around the one state's inclusion or exclusion can achieve similar shifts as well. We are not able to show these tabulations directly or name the state due to the requirement that LEHD samples (or differences across two related samples) contain three or more states. Our internal files record the state breakout. Thus, in some respect, the unconditional results evident in panel B are the perspective taken when one allows for much of immigrant entrepreneurship to be in one location. This can include possibly endogenous flows of immigrants for opportunities, and it may reflect how immigrant entrepreneurship impacts the state's economic dynamics. By contrast, the patterns in panel C of these two tables are observations that control for these overall regional differences. Both perspectives are quite important to consider.

To look a little further behind these results, we also conduct several sample splits in tables 5.6A and 5.6B for estimations with and without conditioning on cohort-PMSA-industry fixed effects. Each coefficient in the tables is from a separate regression. The top row repeats the baseline specification, followed by splits between low- and high-wage firms, low- and high-tech sectors, and then VC backing (non-VC-backed results are not reported since they are so close to the baseline outcomes). We mostly focus on table 5.6B, where several intriguing differences are present. First, variation among low-wage and low-tech groups is generally responsible for our conditional survival relationship. Second, employment growth is generally associated with high-wage and high-tech sectors. Third, payroll growth compared to natives is never present. Fourth, while immigrant-founded ventures are more likely to access VC financing, they do not display stronger outcomes conditional on this financing. This is in many respects a parallel finding to our observation that city-industry choice accounts for much of the observed differences in the general sample. Finally, where employment growth occurs, it is usually about achieving any employment expansion relative to the initial level or reaching the top deciles of an industry, rather than surpassing the threshold of 100 employees (a benchmark that few newly started firms make). This is true for

Table 5.6A   Split sample regressions for estimations with cohort fixed effects

				Traits of surviving firms in the third year				
	Fraction alive in third year (1)	Employment growth relative to firm's average (2)	Payroll growth relative to firm's average (3)	Establishment count growth relative to firm's average (4)	(0,1) third-year employment exceeds initial level (5)	(0,1) third-year employment exceeds 100 workers (6)	(0,1) third-year employment is in top decile of industry (7)	(0,1) received venture capital support (8)
			Estimation with cohort fixed effects and controls for initial traits					
Full sample	−0.00335	0.03130	0.00037	−0.00025	0.01919	0.00159	0.00362	0.00232
	(0.00186)	(0.00301)	(0.00356)	(0.00022)	(0.00241)	(0.00036)	(0.00109)	(0.00020)
Low-wage firms	0.01380	0.00410	−0.00157	−0.00045	0.00320	0.00000	−0.00285	0.00006
	(0.00247)	(0.00394)	(0.00478)	(0.00024)	(0.00331)	(0.00038)	(0.00125)	(0.00006)
High-wage firms	−0.02681	0.06039	0.00316	−0.00023	0.04215	0.00273	0.00848	0.00484
	(0.00280)	(0.00463)	(0.00531)	(0.00038)	(0.00353)	(0.00064)	(0.00180)	(0.00045)
Low-tech firms	−0.00015	0.03169	−0.00220	−0.00018	0.01994	0.00169	0.00271	0.00092
	(0.00193)	(0.00312)	(0.00368)	(0.00023)	(0.00252)	(0.00037)	(0.00112)	(0.00014)
High-tech firms	−0.02383	0.02898	0.02307	−0.00126	0.01794	−0.00021	0.01088	0.00988
	(0.00627)	(0.01116)	(0.01346)	(0.00083)	(0.00822)	(0.00145)	(0.00422)	(0.00160)
VC-backed firms	0.03990	−0.11521	−0.11576	−0.00624	−0.01279	−0.07686	−0.09093	n/a
	(0.03374)	(0.05524)	(0.07042)	(0.01009)	(0.02701)	(0.02287)	(0.04436)	

Notes: See panel B of table 5.5A. Each entry in table is a coefficient from a separate regression with the indicated sample.

Table 5.6B   Split sample regressions for estimations with cohort-PMSA-industry fixed effects

	Fraction alive in third year (1)	Employment growth relative to firm's average (2)	Payroll growth relative to firm's average (3)	Establishment count growth relative to firm's average (4)	(0,1) third-year employment exceeds initial level (5)	(0,1) third-year employment exceeds 100 workers (6)	(0,1) third-year employment is in top decile of industry (7)	(0,1) received venture capital support (8)
							Traits of surviving firms in the third year	
		Estimation with cohort-PMSA-industry fixed effects and controls for initial traits						
Full sample	0.0185	0.0019	-0.0238	-0.0002	0.0066	0.0005	0.0030	0.0005
	(0.0025)	(0.0048)	(0.0057)	(0.0004)	(0.0038)	(0.0006)	(0.0018)	(0.0003)
Low-wage firms	0.0280	-0.0337	-0.0186	-0.0002	-0.0209	0.0003	-0.0027	0.0000
	(0.0037)	(0.0076)	(0.0093)	(0.0005)	(0.0064)	(0.0008)	(0.0026)	(0.0001)
High-wage firms	0.0000	0.0464	-0.0216	-0.0007	0.0370	-0.0005	0.0089	0.0009
	(0.0039)	(0.0076)	(0.0087)	(0.0006)	(0.0058)	(0.0011)	(0.0031)	(0.0007)
Low-tech firms	0.0222	0.0001	-0.0271	-0.0003	0.0062	0.0008	0.0018	0.0002
	(0.0027)	(0.0051)	(0.0060)	(0.0004)	(0.0041)	(0.0007)	(0.0019)	(0.0002)
High-tech firms	-0.0022	0.0258	-0.0084	-0.0001	0.0254	-0.0021	0.0133	0.0023
	(0.0085)	(0.0180)	(0.0221)	(0.0016)	(0.0135)	(0.0027)	(0.0071)	(0.0025)
VC-backed firms	0.0319	-0.2620	-0.3500	0.0101	-0.0709	-0.0311	-0.1061	n/a
	(0.0871)	(0.1549)	(0.1868)	(0.0138)	(0.0792)	(0.0557)	(0.1163)	

Notes: See panel C of table 5.5A. Each entry in table is a coefficient from a separate regression with the indicated sample.

the general and VC-backed samples, suggesting that employment effects due to immigrant entrepreneurship are more likely to come from accumulated contributions of many firms compared to the extreme outcomes of a few high-growth entities that are often emphasized in the popular debate.

In unreported estimations, we also look at the probability of achieving an IPO by 2005, both in the whole sample and among firms backed by VC investors. We do not observe differences for immigrant entrepreneurs in this regard, but we also hesitate to emphasize this result given the early end date of the Compustat Bridge File for our sample.

While this study does not identify why immigrants might choose riskier city-industry cells, some preliminary candidates can be listed. One is that the self-selection process of international migration leads to a pool of foreign-born individuals in the United States who have a greater tolerance for risk and uncertainty than the average US native. This could lead to differences in the distributions of businesses started by the two groups with respect to potential growth outcomes. A second possibility is that immigrants have weaker wage-based options, due to some combination of factors like limited language skills, reduced acceptance of education credentials, spatial isolation in ethnic enclaves, social exclusion, or similar (e.g., Lewis 2013a). These reduced outside options may make immigrants willing to launch a business of any form and venture into riskier domains. A third, more positive, possibility is that the tight social structures for some immigrant groups allow them a group-based capacity to enter into riskier domains and rely on each other, similar to the studies of immigrant entrepreneurial specialization (e.g., Kerr and Mandorff 2015). This comparative advantage could be consistent with benefits of immigrant-generated diversity documented by Ottaviano and Peri (2006), Mazzolari and Neumark (2012), Nathan (2015), and similar. Another possibility, to complete a first and incomplete list, is that illegal immigration and undocumented workers in immigrant-led firms have somehow led us to mismeasure some of the growth/survival properties. We believe the LEHD-based platform provides power going forward to investigate these questions, for example, by taking advantage of the observation of wage histories before and after entrepreneurial spells to understand outside options. Similarly, the detailed information on country of birth can aid analyses of social isolation, group concentration on entrepreneurship in certain industries, and similar features.

Table 5.7 provides our final analysis, which is fairly brief. One goal of this project is to evaluate whether the 2000 Census traits can be incorporated systematically into the immigrant entrepreneurship analysis. Given the policy interest on the age at arrival of immigrants, we choose this dimension and show some initial tabulations in the transition framework of tables 5.4A and 5.4B. We restrict the sample to start-ups that have immigrant founders, and sample sizes require that we combine the final growth outcomes in column (6) to achieving twenty or more employees. The cells now represent the share

Table 5.7   Distributions of child immigrant entrepreneurship

Row headers group by initial employment: (1)	Employment after three or six years:					
	Closed (2)	1–4 empl. (3)	5–9 empl. (4)	10–19 empl. (5)	20+ empl. (6)	Total (7)
A. Share of immigrant entrepreneurs who migrated as children, three-year distribution						
1–4 employees	0.344	**0.366**	0.450	0.462	0.333	0.372
5–9 employees	0.329	0.455	**0.329**	0.438	0.410	0.373
10+ employees	0.493	0.455	0.343	**0.337**	0.503	0.464
Total	0.391	0.379	0.394	0.403	0.469	0.400
B. Share of immigrant entrepreneurs who migrated as children, six-year distribution						
1–4 employees	0.348	**0.379**	0.449	0.426	0.419	0.372
5–9 employees	0.349	0.368	**0.361**	0.478	0.424	0.373
10+ employees	0.484	0.211	0.375	**0.358**	0.504	0.464
Total	0.392	0.368	0.410	0.418	0.477	0.400

Notes: See tables 5.4A and 5.4B. Values document the share of immigrant entrepreneurs in each cell that migrated to the United States by age eighteen. The sample is restricted to immigrant entrepreneurs age twenty to fifty-four in 2000 who are surveyed by the long form of the Decennial Census.

of immigrant founders in the cell who came to the United States as children. At starting employment levels of nine workers or less, immigrants coming to the United States as children are generally associated with better outcomes in terms of lower closure rates and higher representation among the larger size categories. Immigrants coming to the United States as children are also more likely to start larger firms, and among this largest category they tend to be overrepresented among business closures and the firms that achieve the largest employment outcomes. As important, we form a general conclusion from this exercise that researchers will be able to split the samples sufficiently along traits available in the 2000 Census match to explore outcomes associated with different traits of immigrant entrepreneurs. Tabulations of growth by two or more dimensions (e.g., education and age of arrival) will quickly become strained, requiring multivariate regression analysis instead.

5.6   Conclusions and Future Research

The constructed data platform provides new statistics regarding the patterns of business formation by immigrant entrepreneurs and the medium-term success of those businesses. The definition of an entrepreneur used in this study is in many ways dictated by the coverage of the LBD and the LEHD, and hence it is useful to compare our calculations and estimates to those derived from other data sources. Looking back at table 5.1, our results tend to fall in the upper end of the estimates for the immigrant entrepreneur share. There are several factors potentially at work. First, the LEHD data does not identify the actual founders of businesses, and we may be over- or

underinclusive in our definition using the top three initial earners. We have yet to identify a technique to quantify this effect relative to other definitions possible for founders. One of the most feasible comparisons may be business ownership records in the SBO (e.g., Fairlie 2008, 2012). Ownership estimates will be higher than entrepreneurship estimates due to the larger existing stock of native small business owners compared to new firm formation, and so we do not expect this difference to close. But through the study of new entrants captured in the SBO, perhaps the entrepreneurship metrics can be enhanced or their properties better defined.

Beyond immigrant shares of entrepreneurs, rates of entry are more difficult to reconcile due to many alternatives for both the numerator (who is an entrepreneur?) and the denominator (what population are we comparing this to?). Our data have some distinct traits. First, the LEHD only includes employer firms that file UI records, and thus excludes nonemployer firms and self-employment. Addition of these individuals will most likely alter the estimated rates by mainly boosting the numerator. That said, most of the policy focus and academic interest centers on attracting "growth entrepreneurs" that create jobs, in which case a restriction of the analysis to employer firms is actually desired. Second, our denominator focuses on private-sector workers in the LEHD. This denominator could be also more inclusive by incorporating the entire public sector or by being expanded to be all working-age adults in the covered states (e.g., drawn from the Current Population Survey). Using a larger population for the denominator will obviously reduce our measured rate of entrepreneurship. In general, our calculations seek to provide a new, longitudinal view into the patterns of immigrant entrepreneurship and not directly replicate nor necessarily displace findings from earlier studies. Our platform sacrifices on some dimensions (e.g., state coverage, ownership data) and gains on others (e.g., universal samples, longitudinal depth, homogeneity across skills).

While there has been considerable recent interest in immigrant entrepreneurship, as surveyed by Fairlie and Lofstrom (2014), the state of the field in terms of academic studies is still mostly wide open. Even establishing the basic facts has been hampered by the availability of large scale, nationally representative longitudinal data that would capture both firm founders and their firms. Below we describe several promising areas for further research, many of which would benefit from greater data availability from administrative sources.

First, it would be helpful to build a more solid research foundation for quantifying the magnitudes of immigrant contributions to the creation of new entrepreneurial firms in the Science, Technology, Engineering and Mathematics (STEM) fields. Many immigration policies specifically target this area—for example, the longer Optional Practical Training (OPT) period for STEM-degree holders—and much of the concern over encouraging immigrant entrepreneurship focuses here. In doing so, researchers could

utilize the employer-employee panel data developed here and ideally also engage in a more causal analysis of policy changes affecting high-skilled immigration in general and poststudy immigration of STEM graduates from US universities in the first decade of the twenty-first century and thereafter.

Second, a better understanding of how the existing immigrants in the United States can more effectively engage in starting new businesses requires careful study of the choices and policy constraints faced by immigrants in their decisions to build and grow new firms versus being workers in a larger corporation. We also lack a clear picture of how the successful immigrant founders enter the United States, which can be for reasons as diverse as schooling, employment, family reunification, and more. A study of the transitions or the sequence of events explaining entry by immigrant entrepreneurs and the role of policies in allowing/blocking this transition would be a helpful start. For example, a frequent policy misconception is that the H-1B immigrants are responsible for lots of start-ups, and expanding the H-1B cap would boost entrepreneurship. While it might indeed do so over some horizon, we would anticipate a significant time lag because H-1B workers are tied to their sponsoring employers until a green card is approved, often taking six or more years. A specific evaluation could focus on the transitions of H-1B holders into entrepreneurship. While the current data platform can provide reasonable approximations of H-1B holders—for example, isolating immigrant computer programmers age twenty-five to thirty who were born in India via the 2000 Census match—a better scenario would be to gain access to the United States Citizenship and Immigration Services (USCIS) records on H-1B visa holders and match them to the LEHD and other Census Bureau data sources. Other visa categories lend themselves to similar evaluations, including the OPT visas for F-1 students. A less intensive alternative is to link smaller sets of individuals like the H-1B visa lotteries studied by Doran, Gelber, and Isen (2015). Accomplishing such linkages may be difficult, but better-quality data ingredients will result in substantially better advice for policymakers.

Continuing on these themes, it is widely thought that immigrant entrepreneurs contribute disproportionately to innovation and technological advancement in the United States (similar to the more established facts about the role of immigrants for innovation generally). One way to quantify immigrant entrepreneur contributions in the science and innovation arena is the United States Patent and Trademark Office (USPTO) patent data (Hall, Jaffe, and Trajtenberg 2001) matched to the person- and firm-level data sources available from the Census Bureau. Indeed, while there exist recent studies on the effect of high-skilled immigration on US innovation, we lack a systematic evaluation of how the creation of new firms by immigrant founders contributes to the overall pace and direction of US innovation and whether these firms produce different types of innovations compared to native-founded firms (e.g., exploration versus exploitation work). Several studies have made progress on the firm-level matching using

name-matching techniques (e.g., Kerr and Fu 2008; Balasubramanian and Sivadasan 2011; Akcigit and Kerr 2010). These studies typically find record matching to be easier for larger firms than small start-ups. It would be terrific to have a match of LEHD workers to the inventor records in the USPTO database. Such a match would enable detailed studies of technological trajectories for workers (how start-ups relate to the innovative work of their prior employer), provide greater assurance about the quality of the matches overall, and facilitate interesting work on immigration and other related labor market policies (e.g., noncompete clauses).

In a very similar vein, we have also made significant headway toward identifying firms backed by VC investors in the Census Bureau data (e.g., Chemmanur, Krishnan, and Nandy 2011; Puri and Zarutskie 2012). Subsequent work in entrepreneurial finance frequently focuses on individual entrepreneurs and/or their specific VC investors. Individual-level connections of the LEHD workers to these data would provide a very powerful platform for the study of VC-backed entrepreneurial outcomes.

While the current study provides some descriptive analyses for a broader set of geographic areas, a more detailed analysis of the impact of immigrant entrepreneurship on local job growth and economic development is warranted. Feldman and Kogler (2010) and Carlino and Kerr (2015) review the literature on the geography of innovation that has come since Audretsch and Feldman (1996). Glaeser, Kerr, and Kerr (2015) consider the general link of entrepreneurship to city employment growth, and Samila and Sorenson (2011) consider the specific case of VC-backed start-ups. It would be useful to build on this past work to understand the specific case of immigrant entrepreneurship. Many cities and local areas are attempting to leverage immigrant entrepreneurship directly, and we need to know more about the potential efficacy of such efforts and how any stimulus actually accrues. Kerr (2010) finds that ethnic entrepreneurs are particularly important in the reallocation of inventive activity to be near sources of breakthrough innovations and their scaling process (e.g., Duranton 2007); study of these phenomena within the LEHD data family is quite promising. Similarly, we skip in this study the ethnicity and immigration status of hired employees due to data features noted above (e.g., the expansion of firms across LEHD state lines), and such features would be very natural to incorporate in a local growth context given the complete definitions of employee traits.

Appendix

Data Appendix: LEHD, Immigration, and Firms

This appendix explains in more detail how the LEHD data are structured and what types of issues are likely to arise when attempting to follow persons and firms over time within the LEHD. It is meant to provide guidance and

useful suggestions for researchers interested in utilizing these data to study immigration and entrepreneurship at the firm level. Our analysis files for the current study and some previous studies (Kerr and Kerr 2013; Kerr, Kerr, and Lincoln 2015b) are available for researchers within the Census Research Data Center (RDC) network. We can also provide any SAS and Stata programs that have been redacted for any company identification to researchers who do not have access to the Census RDC network, although these files are likely to be of limited use without access to the raw data.

LEHD File Structure and Key Identifiers

The LEHD is available for researchers at the Census Research Data Centers. Access to the LEHD requires an approved project and security clearance. This section gives an outline of the LEHD and is geared toward building a firm sample using the LEHD and auxiliary Census Bureau data sets. The LEHD structure is described in greater technical detail in three Census Bureau documents: McKinney and Vilhuber (2008, 2011) and Abowd et al. (2009). We provide here a short description of the relevant data files and variables for the construction of a firm panel, omitting details of other LEHD structure files and variables for brevity. Prospective users of the LEHD are highly encouraged to review the full technical documentation, as well as any previous studies utilizing the LEHD, since the database has complications and issues for which researchers are building codified and tacit knowledge.

The LEHD data are currently available for research purposes for a total of thirty-one states. All states have signed the Memorandum of Understanding (MOU) to include their data in the LEHD, and once entered into the LEHD they will also retroactively include the data series as far back in time as their state's information and records permit. It is not yet clear when the remaining states will be included in the database. Discussion below includes more details on the partial state coverage and practical concerns that it raises.

The LEHD covers all private companies operating in the United States, and it allows researchers to analyze these companies and their workers at a very detailed level over a long period of time. Firms and their business units are identified in the LEHD by three main variables: SEIN, EIN, and ALPHA. The SEIN is a state tax identifier, the EIN is a federal tax identifier, and ALPHA is the Census Bureau's identifier of overall firms. In addition, as firms can have multiple establishments within a state, the LEHD also provides the SEINUNIT number that corresponds to the SEIN reporting unit (i.e., establishment). These variables uniquely identify a firm and its establishments within a calendar quarter and state. The person identifier (PIK) uniquely identifies a worker across all jobs that the individual holds, is derived from Social Security numbers, and is anonymized to protect the person's identity.

The LEHD consists of several separate files describing the firm, worker, and job that the person holds within the firm. These files are organized sepa-

rately by state, and they have a uniform structure across all states. First, data contained in the employer characteristics file (ECF) and the employment history file (EHF) are essentially provided at the establishment level (i.e., by SEIN and SEINUNIT). As SEINs are only uniquely associated to a specific firm during a given calendar quarter within a given state, creating a firm-level panel requires first assigning each SEIN to a single ALPHA that can then be linked to a single firm entity. This process is discussed in detail below. The person-level information contained in the individual characteristics file (ICF) can then be easily linked to the EHF using the PIK.

The Business Register Bridge (BRB) files are the key for creating firms that combine all multiunit entities into a single company across all states. The BRB consists of two separate files: the BR list and the ECF list.[13] The former contains the full list of EINs belonging to each ALPHA and is organized by year, EIN, state, county, four-digit SIC, and Census File Number (CFN). The ECF list reports all SEINs belonging to each EIN, and each record is identified by SEIN, SEINUNIT, year, and calendar quarter. The dual nature and differing record structures of the BRB files make them somewhat cumbersome to use, and creating a full mapping of SEINs for each company name in our sample requires several data cleaning steps and additional research for unclear cases. These steps are documented in more detail later in this appendix.

The individual-level information contained in the ICF includes person characteristics such as the date of birth, gender, race, place of birth, and citizenship status. Similarly, the firm-level information in the ECF describes location, industry, payroll, and other firm characteristics at the SEIN and SEINUNIT levels. In turn, each job that a person holds in any of the companies covered by the LEHD is recorded in the EHF. The EHF tells for each calendar quarter how much the person earned while employed in each company that they worked for. The EHF (merged with the ICF) is crucial for calculating statistics on the company workforce over time.

The key ICF variables for identifying immigrants are the place of birth (POBST)[14], the indicator for foreign place of birth (POBFIN), and the indicator for ever being an alien (ALIEN). These variables allow us to construct a "country of birth" variable.[15] On the other hand, the LEHD is missing

13. The BRB files are further separated into two vintage files. The older vintage files cover years until 2001, and the newer vintage files cover years 1997–2004. While data for the overlapping years should be mostly identical, there are instances where the linkage information differs between the two files. We prioritize the later vintage information where conflict exists.

14. The LEHD country codes are based on the official country codes used, for example, by the Department of Defense. They require some additional processing due to the fact that countries have changed over time and the LEHD records the country as written down by the person in their application for a Social Security number. For example, Germany can show up as GM, BZ, GC, or GE depending on the time of application. In unclear cases we used the city of birth variable (POBCITY) to resolve conflicts.

15. Kerr, Kerr, and Lincoln (2015b) further aggregates these into ten country groups: (a) China, Hong Kong, and Macao; (b) India, Pakistan, and Bangladesh; (c) other Asian coun-

some information that would be relevant for a study of high-skilled immigration, or indeed for many other economic topics. First, it does not contain information on the job characteristics such as occupation, position held, hours worked, or hourly wage. The employment data do not identify reporting relationships or similar attributes of the corporate hierarchy. Second, the ICF does not report the person's actual education (an imputed version is provided) or their immigrant status (e.g., green card, H-1B visa), or changes in these characteristics. We would ideally like to know the person's initial year of entry into the United States, as well as any changes that have occurred in their immigrant status over time.

Other Census Data Required for Firm Identification

The LEHD can be used (and is perhaps most naturally used) at the SEIN level, which for most companies roughly corresponds to an establishment, and for single-unit firms contains the entire firm. In theory, researchers can construct larger "Census firms" by using the BRB to identify all SEINs that belong to a single ALPHA, and then consider the ALPHA to be the unique company identifier. This approach often works fine if one is primarily interested in looking at the cross section of firms at a specific point in time, but the approach does not work well when one needs to track large, multiunit companies over time. It is true that most large firms have one core ALPHA under which most of their employment falls within any given year, and this identifier is mostly fixed over time. However, a more careful look at the larger, multiunit firms studied in Kerr, Kerr, and Lincoln (2015b) finds corporate events such as mergers, acquisitions, spin-offs, and so forth substantially weaken the exclusive use of ALPHAs for longitudinal analyses. In other words, an ALPHA with a large number of employees may temporarily disappear, switch to a different ALPHA, or lose (or gain) a very large number of employees from one year to the next as a result of corporate restructuring.

This is undesirable if one is interested in describing changes in a firm's workforce related to a specific phenomenon, such as high-skilled immigration. Some of our prior work (Kerr, Kerr, and Lincoln 2015b; Kerr and Kerr 2013) thus pursues the creation of composite firm entities.[16] Such steps require identifying all of the ALPHAs from the company's workforce during the period of study. With that objective, we find the best approach to be to (a) identify the relevant firms by their names as recorded in the LBD/SSEL

---

tries; (d) English-speaking countries; (e) Russia and former member states of the Soviet Union; (f) other European countries; (g) Middle East; (h) Africa; (i) Central and South America; and (j) other countries. A detailed breakdown of the country codes is available from the authors by request.

16. There can also be conceptual concerns. For example, if the research goal is to study whether immigrants are displacing natives in firms, corporate restructurings need to be very carefully considered. Strictly speaking, an acquisition includes the simultaneous hiring of many immigrant and native workers from the acquired firm, which would have substantial effects on the estimated relationship, but this is not conceptually what studies are after. Kerr, Kerr, and Lincoln (2015b) pursues the creation of composite firms to remove these biases.

and (b) use the LBD/SSEL to identify all ALPHAs that are ever associated with the company name. In other words, the company name becomes the unique firm identifier, and this approach works best when building records for large companies present across the full time period. The most reliable process of identifying ALPHAs for target companies involves a careful review of company histories to identify significant mergers, acquisitions, and spin-offs that take place during the period of interest. The review of company histories is likely to result in a long list of such events. In addition, one should conduct additional research around cases where there is, based on the LBD, a very large shift in the firm employment from one year to the next. These cases often turn out to be mass layoffs, firms going out of business, or other normal events in the firm life cycle. In some cases, however, one finds overlooked corporate events that should again be accounted for in the compiling of composite firms.

Researchers should further search the LBD by company name and name variations, including the companies that have merged into the original sample companies. For each company and its acquisitions, one needs to collect all ALPHAs that are contained in the LBD under any of the name variations. For example, if company AAA has acquired companies BBB and CCC, we would find three (or more) separate ALPHAs in the LBD and group these with a composite company identifier that is used in subsequent work. This process produces a unique company name, along with the full list of ALPHAs that should be combined together in the LEHD for each company. Kerr, Kerr, and Lincoln (2015b) further describes the application of this manual process for the study's group of 319 large firms and major patenting firms. These firm lists and associated identifiers are available to researchers with appropriate approvals and data access.

Identification of Firms in the LEHD

Once the full list of company names and the various ALPHAs belonging to each of the companies are obtained from the LBD, the next step is to use the BRB to identify all EINs ever belonging to those ALPHAs. While the BRB uniquely associates EINs to an ALPHA within a calendar quarter and state, EINs are not necessarily stable over time, potentially creating abrupt changes in the firm employment series that will cause measurement error. The same is true for the SEINs that need to be linked to the ALPHAs via the EIN using the ECF list. Indeed McKinney and Vilhuber (2008) note that the EIN and SEIN exist for tax administrative purposes, and warn that no straightforward method exists for linking multiple SEINs into a single firm. Our preferred approach is described in greater detail below.

As LEHD identifiers do not uniquely capture a complete firm entity, we suggest using the cleaned version of a company name from the LBD as a unique firm identifier. As explained above, each of these firms may contain one or more ALPHAs, and almost certainly contain more than one EIN and SEIN. Using the BR list, one can create a list of ALPHAs that are ever

associated with each EIN. Similarly, using the ECF list, one can create for each SEIN the full list of EINs that the SEIN was ever associated with. Combining these two lists provides the full list of SEINs that ever appeared with an ALPHA. Using the groupings of ALPHAs derived from the LBD, one assembles the full list of SEINs that ever belonged to a composite company. At this point, there are a number of SEINs that appear to belong to multiple companies, and this multiplicity requires careful attention. Also, as some of the company restructuring takes place at the establishment level (e.g., the sale of a plant or division), there are also cases where a specific EIN or SEIN belongs to multiple companies during the sample period. These cases require some additional research and data cleaning that is also described below.

Cleaning up EIN and SEIN Associations

The data thus far contain unique SEINs for which the LEHD data needs to be collected. Before proceeding, one needs to check whether each SEIN in the company sample is a unique match or requires special attention. In the latter cases, researchers are best served prioritizing cases for review through the employment levels of SEINs. Reviewing the employment series is also helpful when verifying whether the SEIN overlap is caused by a corporate event that had not been recognized in the initial firm sample construction. The cleaning process at this stage is fairly manual, and may involve searches for company events. Our preferred method in residual, unclear cases is to assign the SEIN based on the number of years that it is associated with companies.

Partial LEHD State Coverage

Appendix table 5A.1 provides a breakdown of the states included in the LEHD by the year in which their data series begins. As the aim of Kerr, Kerr, and Lincoln (2015b) and similar work is to study the evolution of companies and their employment structures, it is problematic if large shifts in company workforce result simply from the fact that new business units are added into the data when a large state enters the LEHD. We often utilize balanced panels of states that begin at one of three points: 1992 (once Florida joins), 1995 (once Texas joins), or 2000 (once the bulk of states enter). Second, one should consider dropping firms for which the included states do not meet a certain threshold in terms of included employment. For example, one may want to exclude finance firms that have much of their employment in New York, even if they have establishments present in covered states, as one cannot reliably represent the firm and its employment patterns using the LEHD states. These coverage ratios can be determined using the LBD.

Merging External Firm-Level Data into the LEHD

There are many firm-level data sources that are of interest for the analysis of firms and immigration. For example, researchers may want to incorporate

Table 5A.1   **Initial year of state inclusion in the LEHD**

Year (1)	States entering LEHD (2)	Cumulative state count (3)
1991 or before	CA, CO, ID, IL, LA, MD, NC, OR, WA, WI	10
1992	FL	11
1993	MT	12
1994		12
1995	HI, NM, RI, TX	16
1996	ME, NJ	18
1997	WV	19
1998	GA, IA, IN, SC, TN, VA	25
1999	UT	26
2000	OK, VT	28
2001		28
2002	AR	29

Notes: The LEHD files for states run through 2008 in the version used in this study. The start year differs by state and is tabulated in this table. Parts of the records for Georgia and Indiana (EHF, ECF) start earlier than 1998.

Labor Condition Application (LCA) data from the Department of Labor (DOL), which is a first step in the application for an H-1B visa from the United States Citizenship and Immigration Services (USCIS). While the USCIS does not release systematic data on the number of H-1B applications and granted visas by firm, microdata on LCAs are available by firm from the DOL.[17] These data contain all LCAs starting from 2001 and include the employer name, address, the number of jobs they wish to fill, and the specific job characteristics (title, occupation, proposed wage, etc.).

The firm names in the LCA database are entered as they appear on the original application. Since the applications are sent separately by different company locations, there may be variations of the company name that need to be dealt with before aggregating the LCAs under the overall company name. For example, 7-Eleven may submit LCAs under slight name variations such as "7-Eleven," "7 Eleven," "7-Eleven Inc," and "7-Eleven Incorporated." Also, the issues related to corporate restructurings resurface here as well. Once the names in the LCA data are cleaned and aggregated, we create a crosswalk for the company names to merge the LCA data into the LEHD data. The aggregated LCAs can then be easily merged into the LEHD by the clean firm name.

A second example is the merger of patent data into the LEHD-LBD to study innovation outcomes. Again, as firm-naming conventions vary between the LBD and the USPTO data, another name-cleaning step is required. Researchers also must aggregate multiple USPTO assignees into parent firms. Finally, for many purposes it would be useful to know more about

17. The data can be retrieved from http://www.flcdatacenter.com/CaseH1B.aspx.

the actual firms than is reported in the LEHD. One example of an external firm-level data set is the Compustat company database. Compustat provides standardized company financials for publicly traded companies. These data are merged via bridge files or crosswalks similar to those described. All bridge files are available to approved Census projects, but should be explicitly requested in initial research proposals to the Census Bureau.

Merging Internal Person-Level Data into the LEHD

While merging the company-specific data into our LEHD platform is relatively easy using the company name, person-specific data matches require the use of the PIK created by the Census Bureau. Three internal data sources can be directly linked to the LEHD: (a) long-form responses from the 1990 and 2000 Decennial Censuses, as well as the related American Housing Surveys, (b) the Current Population Survey (CPS), and (c) the Survey of Income and Program Participation (SIPP). For the purposes of our study, the census and the CPS contain some very relevant information that is not present in the LEHD, including occupation and actual education. As the LEHD contains the CPS person and household identification numbers, as well as the years during which the PIK matches the CPS identifier, it is relatively straightforward to combine CPS data into the LEHD. The CPS does not provide a time series of observations for each respondent, and the occupation is a point-in-time snapshot for each person that does not necessarily correspond to the actual job that the person is doing in the LEHD. Similarly, the census data is easy to link into the LEHD using the PIKs. Household characteristics can be linked to each of the matched persons using the household identifiers that are internal to the Census Bureau data.

Most person-level records in the LEHD have a one-to-one match to the CPS and census. We exclude the few cases where we find multiple CPS or census observations (e.g., in different states). As such persons make up less than 0.01 percent of the sample, including them either multiple times or randomly allocating the information from one of the CPS or census observations makes no difference in terms of the results of the statistical analysis.

Possible LEHD Interfaces

We close this appendix by describing potential future interfaces between the LEHD and other data products. The weakest dimension of the LEHD for entrepreneurship research relates to the identification of business founders/owners, and our approach in this study is not perfect. Indeed, we envision that future work can greatly expand and improve on the current approach by linking in founder/owner data from additional sources. These include, for example, the business ownership data from the Internal Revenue Service (IRS) Form K-1, the Doing Business As (DBA) filings from state secretaries, data on the Legal Form of the Organization (LFO) that could iden-

tify sole proprietorships, and perhaps also the forthcoming longitudinal Survey of Business Owners (SBO-X) as well as the existing 2007 and 2012 cross-sectional SBOs.

More specifically, the IRS collects data on business ownership by US individuals via the K-1 form (http://www.irs.gov/pub/irs-pdf/f1120ssk.pdf). Provided that the Census Bureau could assign each of the K-1 owners a PIK to merge into the LEHD, these tax records would provide one avenue for linking the owner information from the K-1 data to the LEHD. The second potential data effort is related to the state secretaries' corporate information that comes from the DBA registration process (http://www.secstates.com/). Each state collects these data, and they are held by the secretaries of state. The registration data generally contain the name and address of the business and the name and other details of the owner, and could in principle be brought into the Census Bureau for the purpose of assigning of company and person identifiers.

The SBO data also contain details of business owners, including their age, gender, race, and whether they are born in the United States. This level of detail would be sufficient for linking the SBO ownership percentages into the LEHD for better identification of whether the owner is among the persons who are on the company payroll. This of course would not cover the entire LEHD, given the survey design of the SBO, but it would help validate/refine definitions and enable analysis of interesting data on the company's start-up and expansion capital.

As unincorporated businesses are not included in the LEHD, we envision utilizing the Integrated Longitudinal Business Database (ILBD) to also cross-verify the status of the owner. The ILBD can be then used to provide descriptive details of the unincorporated businesses, while the census and American Community Survey (ACS) provide a description of the characteristics of the entrepreneurs/owners.

For employer sole proprietors, our approach of identifying firm founders is likely not satisfactory as the business owner will not be on the payroll. Indeed, some fraction of new employer businesses start as sole proprietorships and may later change their legal form of organization. Recent research using the Kauffman Firm Surveys found that about one in three firms begin as a proprietorship, while almost as many begin as limited liability companies and as corporations. Of course, not all sole proprietorships are among employer firms. Cole (2011) provides data that suggests that the share of firms that start as employer sole proprietorships is smaller than one-third; the average sole proprietorship has 0.6 employees. That provides some guidance as to the error that may be introduced by ignoring the LFO in the LEHD context. We hope that future research continues to bring together these data elements to provide ever sharper metrics for entrepreneurship and the role of immigrants.

Table 5A.2A   LEHD trends using all earners and three-year new firm definition

Year (1)	Count of immigrants in new firms (2)	Count of other immigrant wage workers (3)	Count of natives in new firms (4)	Count of other native wage workers (5)	Total share of immigrants in LEHD workforce (%) (6)	Immigrant rate of employment in new firms (%) (7)	Native rate of employment in new firms (%) (8)	Immigrant share of employment in new firms (%) (9)
			A. LEHD states present by 1992					
1995	203,480	3,036,760	785,550	15,790,590	16.4	6.3	4.7	20.6
1996	223,590	3,293,330	839,190	16,570,810	16.8	6.4	4.8	21.0
1997	238,330	3,624,270	870,560	17,448,260	17.4	6.2	4.8	21.5
1998	254,720	3,901,610	908,040	18,140,570	17.9	6.1	4.8	21.9
1999	272,660	4,155,360	964,550	18,732,440	18.4	6.2	4.9	22.0
2000	289,120	4,435,670	986,950	19,356,100	18.8	6.1	4.9	22.7
2001	299,620	4,665,730	975,540	19,723,010	19.3	6.0	4.7	23.5
2002	299,430	4,735,170	915,030	19,397,310	19.9	5.9	4.5	24.7
2003	309,820	4,753,190	918,170	19,079,330	20.2	6.1	4.6	25.2
2004	326,070	4,847,310	943,290	19,095,630	20.5	6.3	4.7	25.7
2005	332,780	4,940,650	960,510	19,146,990	20.8	6.3	4.8	25.7
2006	319,490	5,004,690	926,830	19,177,940	20.9	6.0	4.6	25.6
2007	275,820	4,755,270	808,390	18,114,770	21.0	5.5	4.3	25.4
2008	226,920	4,419,640	659,940	16,640,660	21.2	4.9	3.8	25.6
			B. LEHD states present by 2000					
2002	441,390	6,934,120	1,707,180	35,645,500	16.5	6.0	4.6	20.5
2003	445,890	6,992,490	1,684,910	35,077,100	16.8	6.0	4.6	20.9
2004	467,010	7,158,580	1,691,990	35,131,620	17.2	6.1	4.6	21.6
2005	480,550	7,321,280	1,706,910	35,252,440	17.4	6.2	4.6	22.0
2006	463,750	7,449,540	1,636,550	35,356,850	17.6	5.9	4.4	22.1
2007	406,860	7,046,720	1,430,920	33,340,470	17.7	5.5	4.1	22.1
2008	337,480	6,546,090	1,188,780	30,801,120	17.7	4.9	3.7	22.1

Notes: Table provides counts and statistics related to employment in new firms by natives and immigrants in the LEHD. Panel A presents statistics for eleven states present in the LEHD by 1992. Panel B presents statistics for twenty-eight states present by 2000. Table 5A.1 lists states. New firms are defined through the LBD as described in the text and retain their new firm status for the first three years of the firm's life. Caution should be exercised with trends from 2006 onward due to declines in match rates for the Business Registry Bridge between the LEHD and LBD late in the sample. Observation counts are approximated per Census Bureau requirements.

Table 5A.2B   LEHD trends using all earners and one-year new firm definition

Year (1)	Count of immigrants in new firms (2)	Count of other immigrant wage workers (3)	Count of natives in new firms (4)	Count of other native wage workers (5)	Total share of immigrants in LEHD workforce (%) (6)	Immigrant rate of employment in new firms (%) (7)	Native rate of employment in new firms (%) (8)	Immigrant share of employment in new firms (%) (9)
			A. LEHD states present by 1992					
1992	44,770	2,138,260	210,120	12,714,680	14.4	2.1	1.6	17.6
1993	51,640	2,333,490	229,890	13,581,600	14.7	2.2	1.7	18.3
1994	59,350	2,523,290	253,490	14,236,840	15.1	2.3	1.7	19.0
1995	69,060	3,171,180	262,370	16,313,760	16.4	2.1	1.6	20.8
1996	78,990	3,437,930	289,720	17,120,320	16.8	2.2	1.7	21.4
1997	79,920	3,782,690	285,110	18,033,710	17.4	2.1	1.6	21.9
1998	89,420	4,066,900	320,520	18,728,090	17.9	2.2	1.7	21.8
1999	96,200	4,331,820	338,820	19,358,170	18.4	2.2	1.7	22.1
2000	99,720	4,625,080	326,510	20,016,540	18.8	2.1	1.6	23.4
2001	97,340	4,867,410	289,260	20,409,280	19.3	2.0	1.4	25.2
2002	98,290	4,936,300	278,210	20,034,130	19.9	2.0	1.4	26.1
2003	107,280	4,955,730	311,420	19,686,080	20.2	2.1	1.6	25.6
2004	113,290	5,060,100	322,450	19,716,470	20.5	2.2	1.6	26.0
2005	108,710	5,164,720	312,650	19,794,860	20.8	2.1	1.6	25.8
2006	89,640	5,234,540	268,900	19,835,880	20.9	1.7	1.3	25.0
2007	69,370	4,961,720	220,510	18,702,570	21.0	1.4	1.2	23.9
2008	53,730	4,592,830	161,870	17,138,720	21.2	1.2	0.9	24.9
			B. LEHD states present by 2000					
2000	145,660	6,725,760	603,720	36,644,220	15.6	2.1	1.6	19.4
2001	142,060	7,105,810	538,650	37,404,240	16.0	2.0	1.4	20.9
2002	143,950	7,231,560	510,020	36,842,660	16.5	2.0	1.4	22.0
2003	145,430	7,292,940	542,420	36,219,590	16.8	2.0	1.5	21.1
2004	163,690	7,461,900	570,290	36,253,320	17.2	2.1	1.5	22.3
2005	159,460	7,642,370	559,920	36,639,440	17.3	2.0	1.5	22.2
2006	131,500	7,781,790	471,920	36,521,490	17.6	1.7	1.3	21.8
2007	100,410	7,353,160	385,290	34,386,140	17.7	1.3	1.1	20.7
2008	79,880	6,803,680	299,580	31,690,320	17.7	1.2	0.9	21.1

Notes: See table 5A.2A. Table adjusts the definition of new firms to apply to the first year of business entry only.

Table 5A.3A   **Immigrant rate of employment in new firms by SIC group**

Year (1)	SIC1 (2)	SIC2 (3)	SIC3 (4)	SIC4 (5)	SIC5 (6)	SIC6 (7)	SIC7 (8)	SIC8 (9)	Low tech (10)	High tech (11)	VC funded (12)
					A. LEHD states present by 1992						
1995	9.4	5.2	3.2	5.2	8.2	4.5	8.4	5.3	6.4	5.4	1.1
1996	9.2	5.3	3.3	5.3	8.2	4.7	8.5	5.6	6.4	5.7	1.4
1997	8.5	5.3	3.4	5.1	8.1	4.8	8.0	5.5	6.2	5.7	1.8
1998	8.4	5.4	3.1	4.8	8.0	5.2	8.0	5.1	6.2	5.4	1.6
1999	8.2	5.5	3.0	5.2	7.8	5.4	8.0	5.2	6.2	6.0	2.0
2000	7.9	5.3	2.8	5.0	7.7	5.4	8.3	5.1	6.0	6.7	2.2
2001	7.5	4.3	2.7	5.3	7.5	5.2	8.3	5.1	6.0	6.5	2.1
2002	7.5	4.5	2.6	5.4	7.6	5.3	7.5	5.0	6.0	5.9	1.7
2003	7.7	3.9	2.9	5.7	7.7	6.1	7.5	5.1	6.2	5.8	1.4
2004	8.0	4.6	2.9	5.9	7.6	6.4	7.5	5.2	6.5	5.1	1.3
2005	8.4	4.7	2.9	5.4	7.6	6.6	7.6	5.0	6.5	4.9	1.9
2006	8.1	4.5	2.6	5.0	7.3	6.2	7.1	4.6	6.2	4.6	1.4
2007	7.4	4.1	2.2	5.4	6.8	5.2	6.3	4.2	5.7	3.9	0.9
2008	6.3	3.2	1.9	4.8	6.2	4.3	5.8	4.1	5.1	3.4	0.3
					B. LEHD states present by 2000						
2002	7.6	4.0	2.9	5.4	7.8	5.0	7.6	4.8	6.0	6.0	1.5
2003	7.5	3.6	3.0	5.4	7.8	5.6	7.5	4.7	6.0	5.7	1.2
2004	7.8	4.0	3.0	5.5	7.7	5.8	7.4	4.9	6.2	5.2	1.2
2005	8.1	4.1	2.9	5.2	7.7	6.1	7.6	4.7	6.3	4.9	1.6
2006	7.7	4.0	2.7	4.8	7.4	5.6	7.2	4.4	6.0	4.7	1.2
2007	7.3	3.6	2.3	5.2	7.0	4.8	6.6	4.1	5.6	4.2	0.8
2008	6.4	2.9	2.2	4.7	6.3	4.1	6.0	3.9	5.1	3.8	0.3

Notes: The one-digit SIC codes cover the following industries. SIC1: metal mining, coal mining, oil & gas extraction, nonmetal minerals (except fuels), general building contractors, heavy construction (excl. building), special trade contractors; SIC2: food & kindred products, tobacco products, textile mill products, apparel & other textile products, lumber & wood products, furniture & fixtures, paper & allied products, printing & publishing, chemicals & allied products, petroleum & coal Products; SIC3: rubber & misc. plastics products, leather & leather products, stone & clay & glass products, primary metal industries, fabricated metal products, industrial machinery & equipment, electronic & other electric equipment, transportation equipment, instruments & related products, misc. manufacturing; SIC4: railroad transportation, local & interurban passenger transit, trucking & warehousing, water transportation, pipelines (except natural gas), transportation services, communications, electric, gas & sanitary services; SIC5: wholesale trade-durable goods, wholesale trade-nondurable goods, building materials & garden supplies, general merchandise stores, food stores, automotive dealers & service stations, apparel & accessories stores, furniture & home furnishings stores, eating & drinking places; SIC6: depository institutions, nondepository institutions, security & commodity brokers, insurance carriers, holdings & other investment offices; SIC7: hotels & other lodgings places, personal services, business services, auto repair & services & parkings, miscellaneous repair services, motion pictures, amusement & recreation services; SIC8: health services, legal services, educational services, social services, museums & botanical & zoological gardeners, membership organizations, services NEC. The high-tech industries are based on the three-digit SIC classification as shown in Hecker (1999).

Table 5A.3B Immigrant share of employment in new firms by SIC group

Year (1)	SIC1 (2)	SIC2 (3)	SIC3 (4)	SIC4 (5)	SIC5 (6)	SIC6 (7)	SIC7 (8)	SIC8 (9)	Low tech (10)	High tech (11)	VC funded (12)
					A. LEHD states present by 1992						
1995	13.7	33.2	27.9	16.8	24.5	14.1	20.0	16.0	20.4	22.9	28.3
1996	14.1	33.8	25.7	16.8	25.7	14.8	21.2	16.6	20.9	22.5	28.4
1997	14.2	32.6	26.7	17.1	26.6	15.3	21.8	16.8	21.3	22.9	24.3
1998	14.9	34.4	26.5	17.5	26.8	16.0	22.4	16.8	21.8	22.8	24.3
1999	15.6	34.9	25.8	19.1	26.8	16.0	22.9	17.0	22.0	22.0	24.2
2000	16.6	33.4	28.0	20.8	26.8	16.4	24.0	17.5	22.5	23.6	23.5
2001	17.3	33.0	28.7	21.4	27.5	16.6	25.8	18.4	23.3	24.6	26.0
2002	18.2	33.6	28.2	23.3	29.7	17.7	26.8	19.4	24.6	25.4	26.1
2003	19.3	31.6	26.3	23.6	30.1	19.1	26.3	20.8	25.2	25.4	27.2
2004	19.6	31.0	25.9	23.1	31.3	19.7	27.0	21.9	25.8	24.8	24.6
2005	19.7	30.9	25.9	23.3	31.8	19.9	26.9	21.9	25.6	24.6	28.5
2006	19.9	30.4	25.5	23.0	31.4	20.1	27.0	22.0	25.7	24.7	26.9
2007	19.9	30.4	25.7	21.0	31.1	19.6	26.7	22.1	25.5	24.7	26.3
2008	19.6	28.9	24.6	21.6	30.8	19.4	27.0	22.8	25.7	24.4	28.1
					B. LEHD states present by 2000						
2002	16.1	25.3	22.1	17.3	25.5	14.5	22.9	15.9	20.4	21.9	22.7
2003	16.6	23.7	21.5	16.9	26.3	15.6	22.6	16.8	20.8	22.0	23.4
2004	17.0	24.3	21.8	17.4	27.1	16.3	23.4	18.0	21.6	21.5	21.0
2005	17.1	24.7	21.6	18.3	27.9	16.6	23.7	18.2	22.0	21.7	24.6
2006	17.2	24.7	20.9	18.8	28.0	16.8	24.1	18.3	22.1	21.5	23.7
2007	17.5	24.4	20.5	18.4	28.3	16.5	24.1	18.4	22.2	21.5	23.1
2008	17.5	23.3	20.2	18.3	27.9	16.4	24.0	18.8	22.2	21.2	25.3

Notes: See appendix table 5A.3A.

Table 5A.4   LEHD comparisons of immigrant and native employees in new firms

| Trait | Immigrants | | Natives | | Immigrant/ native ratio | Difference p-value |
(1)	Mean (2)	SD (3)	Mean (4)	SD (5)	(6)	(7)
A. Traits of initial employees in new firms, full sample						
Age	37.6	9.2	36.2	9.9	1.037	0.000
Male	61.8	48.6	57.7	49.4	1.071	0.000
Average quarterly earnings in year before founding	$7,335	$25,140	$9,028	$29,878	0.812	0.000
Average quarterly earnings in year of founding	$9,305	$155,301	$10,236	$31,949	0.909	0.000
Average quarterly earnings in year after founding	$10,199	$23,236	$11,339	$33,915	0.899	0.000
N	1,605,970		5,700,570			
B. Panel A for firms backed by venture capital						
Age	34.8	8.2	34.6	8.9	1.008	0.000
Male	0.71	0.46	0.64	0.48	1.100	0.000
Average quarterly earnings in year before founding	$21,160	$102,666	$19,516	$134,875	1.084	0.089
Average quarterly earnings in year of founding	$20,137	$20,603	$18,525	$21,456	1.087	0.000
Average quarterly earnings in year after founding	$24,957	$23,524	$23,503	$28,566	1.062	0.000
N	18,650		48,910			

Notes: Descriptive traits are developed from LEHD.

Table 5A.5   2000 census comparisons of immigrant and native employees in new firms

| | Immigrants | | Natives | | Immigrant/ | Difference |
| Trait | Mean | SD | Mean | SD | native ratio | p-value |
(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Traits of employees in new firms, full sample					
Age	38.1	9.2	36.9	10.0	1.032	0.000
Male	60.0%	49.0%	56.5%	49.6%	1.064	0.000
Married	66.4%	47.3%	53.7%	49.9%	1.235	0.000
Number of children	1.40	1.39	0.98	1.17	1.429	0.000
Bachelor's education and higher	25.2%	43.4%	22.7%	41.9%	1.109	0.000
Home owner	56.6%	49.6%	65.3%	47.6%	0.868	0.000
Home value (max = $1 million)	$188,172	$169,972	$163,722	$152,863	1.149	0.000
Implied rental value	$165,472	$84,305	$154,054	$88,127	1.074	0.000
Move-in date	1994.2	5.7	1992.7	7.7	1.001	0.000
Household income (max = $2.5 million)	$65,737	$74,124	$69,286	$72,430	0.949	0.000
Year of arrival to United States	1984.0	9.7				

Notes: Descriptive traits are developed from 2000 Decennial Census match to LEHD. Implied rental value of dwelling is twenty times monthly rent.

Table 5A.6A Table 5.5A with immigrant share of all initial employees

	Fraction alive in third year (1)	Traits of surviving firms in the third year					
		Employment growth relative to firm's average (2)	Payroll growth relative to firm's average (3)	Establishment count growth relative to firm's average (4)	(0,1) third-year employment exceeds initial level (5)	(0,1) third-year employment exceeds 100 workers (6)	(0,1) third-year employment is in top decile of industry (7)
A. Summary statistics							
Mean	0.6418	0.2726	0.2449	0.0033	0.5133	0.0086	0.0848
SD	0.4866	0.6823	0.7216	0.0567	0.4998	0.0921	0.2787
Observations	535,580	329,260	329,260	329,260	329,260	329,260	329,260
B. Estimation with cohort fixed effects and controls for initial traits							
Firm's immigrant employee share	-0.01017	0.03382	-0.00062	-0.00047	0.01913	0.00479	0.00881
	(0.00182)	(0.00295)	(0.00350)	(0.00019)	(0.00238)	(0.00032)	(0.00097)
ß/mean of DV	-0.016	0.124	-0.003	-0.142	0.037	0.557	0.104
C. Estimation with cohort-PMSA-industry fixed effects and controls for initial traits							
Firm's immigrant employee share	0.01753	0.00298	-0.02772	-0.00039	0.00629	0.00454	0.01103
	(0.00244)	(0.00479)	(0.00566)	(0.00033)	(0.00383)	(0.00054)	(0.00167)
ß/mean of DV	0.027	0.011	-0.113	-0.118	0.012	0.528	0.130

Notes: See table 5.5A.

Table 5A.6B Table 5.5B with immigrant share of all initial employees

	Fraction alive in sixth year (1)	Traits of surviving firms in the sixth year					
		Employment growth relative to firm's average (2)	Payroll growth relative to firm's average (3)	Establishment count growth relative to firm's average (4)	(0,1) sixth-year employment exceeds initial level (5)	(0,1) sixth-year employment exceeds 100 workers (6)	(0,1) sixth-year employment is in top decile of industry (7)
			A. Summary statistics				
Mean	0.4350	0.3097	0.3643	0.0034	0.5362	0.0102	0.0842
SD	0.4958	0.7403	0.7984	0.0591	0.4987	0.1003	0.2776
Observations	535,580	233,000	233,000	233,000	233,000	233,000	233,000
		B. Estimation with cohort fixed effects and controls for initial traits					
Firm's immigrant employee share	-0.02096	0.05274	0.01198	-0.00012	0.02953	0.00552	0.00967
	(0.00183)	(0.00386)	(0.00461)	(0.00023)	(0.00286)	(0.00043)	(0.00122)
β/US mean	-0.048	0.170	0.033	-0.035	0.055	0.541	0.115
		C. Estimation with cohort-PMSA-industry fixed effects and controls for initial traits					
Firm's immigrant employee share	0.01591	0.01240	-0.02835	-0.00039	0.01319	0.00484	0.00779
	(0.00236)	(0.00663)	(0.00789)	(0.00038)	(0.00490)	(0.00078)	(0.00228)
β/US mean	0.037	0.040	-0.078	-0.115	0.025	0.475	0.093

Notes: See table 5.5B.

References

Abowd, J.M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, 149–230. Chicago: University of Chicago Press.

Akcigit, U., and W. Kerr. 2010. "Growth through Heterogeneous Innovations." NBER Working Paper no. 16443, Cambridge, MA.

Akee, R. K. Q., D. A. Jaeger, and K. Tatsiramos. 2007. "The Persistence of Self-Employment across Borders: New Evidence on Legal Immigrants to the United States." IZA Discussion Papers no. 3250, Institute for the Study of Labor.

Anderson, S., and M. Platzer. 2006. "American Made: The Impact of Immigrant Entrepreneurs and Professionals on US Competitiveness." National Venture Capital Association Report. http://www.contentfirst.com/AmericanMade_study.pdf.

Andersson, F., S. Burgess, and J. Lane. 2012. "Do as the Neighbors Do: The Impact of Social Networks on Immigrant Employment." IZA Discussion Papers no. 4423, Institute for the Study of Labor.

Andersson, F., M. Garca-Pérez, J. Haltiwanger, K. McCue, and S. Sanders. 2014. "Workplace Concentration of Immigrants." *Demography* 51 (6): 2281–306.

Åslund, O., L. Hensvik, and O. Skans. 2014. "Seeking Similarity: How Immigrants and Natives Manage in the Labor Market." *Journal of Labor Economics* 32 (3): 405–41.

Audretsch, D., and M. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86:630–40.

Balasubramanian, N., and J. Sivadasan. 2011. "What Happens When Firms Patent? New Evidence from US Manufacturing Census Data." *Review of Economics and Statistics* 93 (1): 126–46.

Bengtsson, O., and D. Hsu. 2015. "Ethnic Matching in the US Venture Capital Market.*" Journal of Business Venturing* 30 (2): 338–54.

Blau, F., and L. Kahn. 2016. "Immigration and Gender Roles: Assimilation vs. Culture." NBER Working Paper no. 21756, Cambridge, MA.

Borjas, G. 1986. "The Self-Employment Experience of Immigrants." *Journal of Human Resources* 21:487–506.

Borjas, G., and K. Doran. 2012. "The Collapse of the Soviet Union and the Productivity of American Mathematicians." *Quarterly Journal of Economics* 127 (3): 1143–203.

Carlino, G., and W. R. Kerr. 2015. "Agglomeration and Innovation." In *Handbook of Urban and Regional Economics*, vol. 5, edited by G. Duranton, V. Henderson, and W. Strange, 349–404. Amsterdam: North Holland.

Chemmanur, T., K. Krishnan, and D. Nandy. 2011. "How Does Venture Capital Financing Improve Efficiency in Private Firms? A Look beneath the Surface." *Review of Financial Studies* 24 (12): 4037–90.

Chung, W., and A. Kalnins. 2006. "Social Capital, Geography, and the Survival: Gujarati Immigrant Entrepreneurs in the US Lodging Industry." *Management Science* 52 (2): 233–47.

Clark, K., and S. Drinkwater. 2000. "Pushed Out or Pulled In? Self-Employment among Ethnic Minorities in England and Wales." *Labour Economics* 7 (5): 603–28.

———. 2006. "Changing Patterns of Ethnic Minority Self-Employment in Britain: Evidence from Census Microdata." IZA Discussion Papers no. 2495, Institute for the Study of Labor.

Cole, R. 2011. "How Do Firms Choose Legal Form of Organization?" SBA Release

no. 383, United States Small Business Administration. https://www.sba.gov/sites/default/files/files/rs383tot.pdf.

Davis, S., J. Haltiwanger, and S. Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: MIT Press.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives* 28 (3): 3–24.

Doran, K., A. Gelber, and A. Isen. 2015. "The Effects of High-Skill Immigration on Firms: Evidence from H-1B Visa Lotteries." NBER Working Paper no. 20668, Cambridge, MA.

Duranton, G. 2007. "Urban Evolutions: The Fast, the Slow, and the Still." *American Economic Review* 97:197–221.

Fairlie, R. W. 2008. "Estimating the Contribution of Immigrant Business Owners to the US Economy." Small Business Research Summary no. 332, Small Business Administration, Washington, DC.

———. 2012. "Immigrant Entrepreneurs and Small Business Owners, and their Access to Financial Capital." US Small Business Administration Report. https://www.sba.gov/sites/default/files/rs396tot.pdf.

———. 2013. "Minority and Immigrant Entrepreneurs: Access to Financial Capital." In *International Handbook on the Economics of Migration*, edited by A. F. Constant and K. F. Zimmermann. Cheltenham, UK: Edward Elgar.

Fairlie, R. W., and M. Lofstrom. 2014. "Immigration and Entrepreneurship." In *Handbook on the Economics of International Migration*, edited by B. R. Chiswick and P. W. Miller. Amsterdam: Elsevier.

Fairlie, R. W., and B. D. Meyer. 2003. "The Effect of Immigration on Native Self-Employment." *Journal of Labor Economics* 21 (3): 619–50.

Fairlie, R. W., A. Morelix, E. J. Reddy, and J. Russell. 2015. "The Kauffman Index 2015." Kauffman Foundation Report. http://www.kauffman.org/~/media/kauffman_org/research%20reports%20and%20covers/2015/05/kauffman_index_startup_activity_national_trends_2015.pdf.

Fairlie, R. W., J. Zissimopoulos, and H. A. Krashinsky. 2010. "The International Asian Business Success Story: A Comparison of Chinese, Indian, and Other Asian Businesses in the United States, Canada, and United Kingdom." In *International Differences in Entrepreneurship*, edited by J. Lerner and A. Schoar, 179–208. Chicago: University of Chicago Press.

Feldman, M., and D. Kogler. 2010. "Stylized Facts in the Geography of Innovation." In *Handbook of the Economics of Innovation*, vol. 1, edited by B. Hall and N. Rosenberg, 381–410. Oxford: Elsevier.

Garca-Pérez, M. 2012. "A Matching Model on the Use of Immigrant Social Networks and Referral Hiring." Economics Faculty Working Paper no. 21, Saint Cloud State University. http://repository.stcloudstate.edu/cgi/viewcontent.cgi?article=1017&context=econ_wps.

Glaeser, E., S. P. Kerr, and W. R. Kerr. 2015. "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines." *Review of Economics and Statistics* 97 (2): 498–520.

Hall, B., A. Jaffe, and M. Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper no. 8498, Cambridge, MA.

Haltiwanger, J., R. Jarmin, and J. Miranda. 2013. "Who Creates Jobs? Small vs. Large vs. Young." *Review of Economics and Statistics* 95 (2): 347–61.

Hart, D. M., and Z. J. Acs. 2011. "High-Tech Immigrant Entrepreneurship in the United States." *Economic Development Quarterly* 25 (2): 116–29.

Hecker, D. 1999. "High-Technology Employment: A Broader View." *Monthly Labor Review*, June. http://www.bls.gov/opub/mlr/1999/06/art3full.pdf .

Hegde, D., and J. Tumlinson. 2014. "Does Social Proximity Enhance Business Relationships? Theory and Evidence from Ethnicity's Role in US Venture Capital." *Management Science* 60 (9): 2355–80.

Hunt, J. 2011. "Which Immigrants are Most Innovative and Entrepreneurial? Distinctions by Entry Visa." *Journal of Labor Economics* 29 (3): 417–57.

———. 2015. "Are Immigrants the Most Skilled US Computer and Engineering Workers?" *Journal of Labor Economics* 33 (S1): S39–77.

Hunt, J., and M. Gauthier-Loiselle. 2010. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2 (2): 31–56.

Kerr, S. P., and W. R. Kerr. 2013. "Immigration and Employer Transitions for STEM Workers." *American Economic Review Papers and Proceedings* 103 (3): 193–97.

Kerr, S. P., W. R. Kerr, and W. F. Lincoln. 2015a. "Firms and the Economics of Skilled Immigration." In *Innovation Policy and the Economy*, vol. 15, edited by W. R. Kerr, J. Lerner, and S. Stern, 115–52. Chicago: University of Chicago Press.

———. 2015b. "Skilled Immigration and the Employment Structures of US Firms." *Journal of Labor Economics* 33 (S1): S109–45.

Kerr, S. P., W. R. Kerr, and R. Nanda. 2015. "House Money and Entrepreneurship." NBER Working Paper no. 21458, Cambridge, MA.

Kerr, W. R. 2010. "Breakthrough Inventions and Migrating Clusters of Innovation." *Journal of Urban Economics* 67 (1): 46–60.

———. 2013. "US High-Skilled Immigration, Innovation, and Entrepreneurship: Empirical Approaches and Evidence." NBER Working Paper no. 19377, Cambridge, MA.

Kerr, W. R., and S. Fu. 2008. "The Survey of Industrial R&D—Patent Database Link Project." *Journal of Technology Transfer* 33 (2): 173–86.

Kerr, W. R., and W. F. Lincoln. 2010. "The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention." *Journal of Labor Economics* 28 (3): 473–508.

Kerr, W. R., and M. Mandorff. 2015. "Social Networks, Ethnicity, and Entrepreneurship." NBER Working Paper no. 21597, Cambridge, MA.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf. 2014. "Entrepreneurship as Experimentation." *Journal of Economic Perspectives* 28 (3) 25–48.

Lewis, E. 2013a. "Immigrant Native Substitutability: The Role of Language." In *Immigration, Poverty, and Socioeconomic Inequality*, edited by D. Card and S. Raphael. New York: Russell Sage Foundation.

Lewis, E. 2013b. "Immigration and Production Technology." *Annual Review of Economics* 5:165–91.

Lewis, E., and G. Peri. 2015. "Immigration and the Economy of Cities and Regions." In *Handbook of Urban and Regional Economics*, vol. 5, edited by G. Duranton, V. Henderson, and W. Strange. Amsterdam: North Holland.

Lofstrom, M. 2002. "Labor Market Assimilation and the Self-Employment Decision of Immigrant Entrepreneurs." *Journal of Population Economics* 15 (1): 83–114.

Lofstrom, M., T. Bates, and S. Parker. 2014. "Why Are Some People More Likely to Become Small-Businesses Owners than Others: Entrepreneurship Entry and Industry-Specific Barriers." *Journal of Business Venturing* 29 (2): 232–51.

Mazzolari, F., and D. Neumark. 2012. "Immigration and Product Diversity." *Journal of Population Economics* 25 (3): 1107–37.

McKinney, K., and L. Vilhuber. 2008. "Longitudinal Employer-Household Dynamics. LEHD Infrastructure Files in the Census RDC-Overview: Revision 219." US Census Bureau.

———. 2011. "LEHD Infrastructure Files in the Census RDC: Overview of S2004 Snapshot." CES Working Paper no. 11-13, Center for Economic Studies, US Census Bureau.

Monti, D. J., L. Smith-Doerr, and J. MacQuaid. 2007. "Immigrant Entrepreneurs in the Massachusetts Biotechnology Industry." Immigrant Learning Center, Boston, MA. http://www.ilctr.org/wp-content/uploads/2011/08/immigrants_in _biotechnology-updated.pdf.

Moser, P., A. Voena, and F. Waldinger. 2014. "German Jewish Emigres and US Invention." *American Economic Review* 104 (10): 3222–55.

Nathan, M. 2015. "Same Difference? Minority Ethnic Inventors, Diversity and Innovation in the UK." *Journal of Economic Geography* 15 (1): 129–68.

Ottaviano, G., and G. Peri. 2006. "The Economic Value of Cultural Diversity: Evidence from US Cities." *Journal of Economic Geography* 6 (1): 9–44.

Patel, K., and F. Vella. 2013. "Immigrant Networks and their Implications for Occupational Choice and Wages." *Review of Economics and Statistics* 95 (4): 1249–77.

Peri, G. 2007. "Higher Education, Innovation and Growth." In *Education and Training in Europe*, edited by G. Brunello, P. Garibaldi, and E. Wasmer, 56–70. Oxford: Oxford University Press.

Peri, G., K. Shih, and C. Sparber. 2015. "STEM Workers, H-1B Visas and Productivity in US Cities." *Journal of Labor Economics* 33 (S1): S225–55.

Puri, M., and R. Zarutskie. 2012. "On the Lifecycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms." *Journal of Finance* 67 (6): 2247–93.

Reynolds, P. D., and R. T. Curtin. 2007. "Business Creation in the United States in 2006: Panel Study of Entrepreneurial Dynamics II Initial Report." Boston: nowPublications.

Robb, A., E. J. Reedy, J. Ballou, D. DesRoches, F. Potter, and Z. Zhao. 2010. "An Overview of the Kauffman Firm Survey: Results from the 2004 to 2008 Data." Kaufmann Foundation Report, May. http://www.kauffman.org/~/media /kauffman_org/research%20reports%20and%20covers/2010/05/kfs_2010_report .pdf.

Samila, S., and O. Sorenson. 2011. "Venture Capital, Entrepreneurship and Economic Growth." *Review of Economics and Statistics* 93:338–49.

Saxenian, A. 1999. *Silicon Valley's New Immigrant Entrepreneurs*. San Francisco: Public Policy Institute of California.
———. 2002. "Silicon Valley's New Immigrant High-Growth Entrepreneurs." *Economic Development Quarterly* 16:20–31.

Schuetze, H. J., and H. Antecol. 2007. "Immigration, Entrepreneurship and the Venture Start-Up Process. The Life Cycle of Entrepreneurial Ventures." In *International Handbook Series on Entrepreneurship*, vol. 3, edited by S. Parker. New York: Springer.

Stephan, P., and S. Levin. 2001. "Exceptional Contributions to US Science by the Foreign-Born and Foreign-Educated." *Population Research and Policy Review* 20 (1): 59–79.

Unger, J. M., A. Rauch, M. Frese, and N. Rosenbusch. 2011. "Human Capital and Entrepreneurial Success: A Meta-Analytical Review." *Journal of Business Venturing* 26 (3): 341–58.

Van der Sluis, J., M. van Praag, and W. Vijverberg. 2008. "Education and Entrepreneurship Selection and Performance: A Review of the Empirical Literature." *Journal of Economic Surveys* 22 (5): 795–841.

Wadhwa, V., A. Saxenian, B. Rissing, and G. Gereffi. 2007. "America's New Immigrant Entrepreneurs." Duke Science, Technology, and Innovation Paper no. 23, Duke University.